SSM - Population Health 8 (2019) 100397

Contents lists available at ScienceDirect

SSM - Population Health

journal homepage: www.elsevier.com/locate/ssmph

Article

Urban developments and health: Evidence from the distributional analysis of biomarkers in China

Toshiaki Aizawa

Department of Economics and Related Studies, University of York, York, YO10 5DD, UK

ARTICLE INFO

Keywords:
Urbanicity
Biomarkers
Distributional decomposition
China*JEL classification:*
I12 I15

ABSTRACT

This study explores the association between urban developments and health in China, a country that has experienced unprecedented economic growth and consequential rapid urbanisation over the last few decades. Exploiting the Chinese Health and Nutrition Survey, this study analyses the effect of these urban developments on the entire distribution of 11 objectively measured health outcomes related to non-communicable diseases. Quantification of the effects on health distribution is achieved by estimating health distribution in a counterfactual situation in which every individual is exposed to the minimum level of urban development. In decomposing the estimated effect into the part attributable to the observed path through which urban developments change observed health-related behaviours (behavioural effect), and the remaining part which cannot be attributable to this observed path (non-behavioural effect), this study sheds light on the mechanisms underlying how urban developments are associated with health outcomes. The results indicate that urban developments are negatively associated in this regard, especially with health outcomes related to body lipids such as triglycerides and cholesterols, blood pressure and kidney-related biomarkers. Furthermore, the results provide strong evidence of heterogeneity in the degrees of association across the distribution.

1. Introduction

1.1. Background

Rapid urbanisation has become one of the salient features of economic development in countries around the globe, particularly in Asia. Whereas 30 years ago, when fewer than one-third of Asians lived in urban areas, today it is almost every second person. This trend in urbanisation is predicted to continue into the foreseeable future, and United Nations (2014) expects that by 2050, nearly two-thirds of Asia's population will live in a built-up city environment. Implications for health have been discussed extensively; for instance, urbanisation could lead to benefits in terms of health service accessibility and a more stable supply network (World Health Organization, 2016), albeit, at the same time, it could have deleterious health effects, for example due to environmental pollution, urban crowding or slum development (Leon, 2008; Mutatkar, 1995; Chen, 2007). It is also well-documented that urbanisation and industrialisation lead to changes in the living environment, the lifestyles of residents and the spread and prevalence of common diseases (Van de Poel, O'Donnell, & van Doorslaer, 2007, 2009; van de Poel, O'Donnell, & van Doorslaer, 2012; Miao & Wu, 2016). Dye (2008) and Deaton (2013) argue that, historically, urbanisation has shifted the burden of illness from acute childhood infections to chronic non-communicable diseases (NCDs)[1] affecting adults. Its overall impact on public health is therefore complex and multi-factorial (Galea & Vlahov, 2005; Moore, Gould, & Keary, 2003).

Over the last few decades, China has experienced strong economic growth and – consequentially – rapid urbanisation. According to United Nations (2014), up to 1980, only one out of five people lived in an urban area, whilst in 1998, this figure had risen to one third, and in 2011, more than half of the population had taken to living in city locations. Historically, the expansion of urban areas was concentrated along the eastern coast, but rapid growth in the past decade has happened in inland provinces as well (Gong et al., 2012). China's recent epidemiological transition is well-characterised by an increase in life expectancy and a shift in the burden of diseases to NCDs (Li, Wang, Zhang, Xiao, & Dixon, 2012; Liu, Yang, Zeng, Horton, & Chen, 2013), in that the lifespan has increased from 40.1 years in 1950 to 76.0 years in 2011 (Riley, 2004). The causes of mortality and morbidity are shifting to resemble those found in high-income countries (Dans et al., 2011), exemplified by the fact that out of 8.3 million deaths per year in China, 7.0 million are attributable to NCDs, with strokes, ischaemic heart

E-mail address: ta812@york.ac.uk.

[1] Non-communicable diseases are also called 'chronic' diseases. Cancers, heart disease, stroke, diabetes and chronic kidney disease are typical examples of non-communicable diseases.

https://doi.org/10.1016/j.ssmph.2019.100397
Received 20 August 2018; Received in revised form 7 February 2019; Accepted 2 April 2019

disease, cancers and chronic obstructive pulmonary disease being the top causes of premature death (Yang et al., 2013).

Despite a large number of studies on urbanisation and health in China, understanding how it affects objective health outcomes is still restricted and often limited to a self-assessed health condition and some anthropometric measures such as body mass index. This study focuses on blood-based biomarkers closely related to NCDs. Recently, in health economics, the usefulness of biomarkers has been acknowledged, and the literature on the subject is growing steadily (Benzeval, Kumari, & Jones, 2016). However, most of the empirical research on objective health indicators has targeted high-income countries, and so unfortunately evidence in low- and middle-income nations is in far poorer supply.

The fundamental goal of this research is to quantify the effect of urban developments in China on the distribution of 11 objectively measured health outcomes related to NCDs. Quantification of the effects on health distribution is achieved by estimating a counter-factual health distribution in a hypothetical situation in which every individual is exposed to the minimum urban development level. The estimated effect of urban developments captures the *net* effect, which also reflects individual preference. In this paper, we do not intend to estimate the causal impact of urban developments, as when related to health in a strict sense it is hard to define conceptually, and the net effects of urban developments seem more relevant from the viewpoint of urban planning. One of the factors making it difficult to define causal impacts is that urban developments themselves do not always affect health directly. The pathways by which urban developments affect health – in most cases – occur through the intermediation of changes in our behaviours such as lifestyle, demographic characteristics and occupation. This research attempts to disentangle the mechanism by which urban developments influence health, by decomposing the estimated effect into the effect associated with the observable path through which urban developments affect individual behaviours (behavioural effects), and the remaining part that cannot be attributable to this observable path (non-behavioural effects) (Fig. 1). Estimating these effects will help understand better the underlying mechanism behind how urban developments are related to health, which in turn will contribute to building up policy implications. For example, if we find a large and adverse behavioural effect, it implies that the adverse effects of urban developments on health are driven largely by the observed changes in health-related behaviours, such as lifestyles, demographic characteristics, living standards and/or occupation. In such a case, enhancing healthy lifestyles and living environments in urban cities could mitigate the adverse effect of urban development on health.

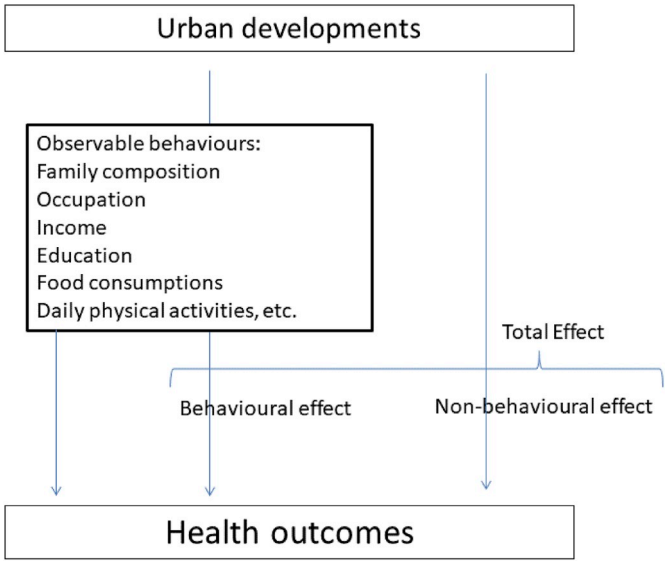

Fig. 1. The urban environment and health.

Recently, more attention has been paid to the entire distribution of health beyond the mean. This study concentrates on this point and estimates the effect of urban developments on the various quantiles of health distribution. The focus on distribution beyond the mean is especially important for the case of health biomarkers, because clinical concerns often lie in the tails of their distribution (Carrieri & Jones, 2017). Distributional analysis can detect the non-linearities of the associations between health and its determinants besides the heterogeneous influences of its determinants across health distribution. Furthermore, focusing on the distribution, we can explore changes in the shape of the distribution as well as any shift in its location. For example, if the effect of urban developments is not constant across the distribution, and it has a larger effect at the right tail, then it suggests that they make the health outcome distribute with a longer right tail. Hence, the distributional analysis undertaken herein helps clarify how the urban environments affect the entire distribution of health and its heterogeneous impact across distribution.

Finally, one of the key features of this study is the use of the continuous measurement of urbanicity. The distinction between urbanisation and urbanicity is discussed by Vlahov and Galea (2002), according to whom "*urbanization refers to change in size, density, and heterogeneity of cities. Urbanicity refers to the impact of living in urban areas at a given time*".[2] Vlahov and Galea (2002) also argue that urbanicity is a more immediate means of studying the unique features of urban areas and their associations with health. The use of continuous measurement can capture the important heterogeneities within and across cities/villages (e.g. the degree of infrastructure development) that are not reflected in the rural/urban dichotomy. Employing the scalar index of urbanicity derived by Jones-Smith and Popkin (2010), we take into account heterogeneities on the level of urban developments that exist within and across cities/villages.

1.2. Related literature

The growing body of evidence suggests that individual health is related closely to urban developments in China (e.g. Chen, Liu, Zhu, & Li, 2017; Fang, Chen, & Rizzo, 2009; Li et al., 2016; Van de Poel, O'Donnell, & van Doorslaer, 2009). Strong economic growth responsible for increasing industrial and urban developments has resulted in an increase in the environmental risks to health, particularly in terms of air, water and soil pollution (Shao, Tang, Zhang, & Li, 2006; Zhang et al., 2010). Substantial changes in lifestyles among the Chinese population have also been reported. As observed across the world, changes in the labour profile have also been witnessed in urban areas, in that labour-intensive industries have been replaced with service industries, leading to predominantly sedentary work styles (Popkin, 2001; Bell, Ge, & Popkin, 2001, 2002). A decline in daily physical activities has also been reported (Monda et al., 2007, 2008), and the Westernisation of diets in China is rapidly becoming an issue with the advent of more animal and partially hydrogenated fats and less fibre (Popkin & Du, 2003; Guo, Mroz, Popkin, & Zhai, 2000; Drewnowski, 2000; Du, Lu, Zhai, & Popkin, 2002, 2004). These dietary changes, especially in terms of the rise in fat intake in urban areas, have been documented (Chen, 1994; Lukman, Dye, & Blundell, 1998; Tian et al., 1995). In addition, economic growth has changed the transportation system in the country, and a growing number of people have become reliant on automobiles, thus contributing to a marked drop in regular exercise (Bell, Ge, & Popkin, 2002). These changes have become the major causes of the increased risk of NCDs, which in many cases require lifetime treatment and are thus costly for both individuals and the country. In this regard, Zhu, Ioannidis, Li, Jones, and Martin (2011) report higher incidents of NCDs in larger cities.

[2] pp.S4–S5 in Vlahov and Galea (2002).

The paper most relevant to this study is Yan et al. (2012), who discuss the relationships between health and income along with relationships between health and urbanicity.[3] Using biomarker data, the authors implement a logistic regression analysis of the probability of each biomarker exceeding a certain respective threshold and estimate the odds ratios across the discrete levels of urbanicity.[4] However, dichotomising continuous biomarkers inevitably loses very important variations in the data, in that it prevents us from discussing heterogeneities among people who are categorised as having a health problem. It is quite likely, for example, that obese people with a body mass index of over 40 have greater health risks than those with a body mass index of, say, 30, which is a commonly used criterion for obesity. Similarly, it seems unlikely that the risk of having a body mass index of 29.9 will be very much different from having one of 30.1 (Aizawa, 2019). As thresholds and standard values for biomarkers are not always consistent across studies and countries, losing important distributional information through the process of dichotomisation can be very costly. We complement the initial findings of Yan et al. (2012) with evidence from a distributional analysis, which detects important distributional heterogeneities in the effects of urban developments that cannot be elucidated otherwise by analyses with dichotomised health outcomes or the discrete urbanicity scale.

2. Data

2.1. Chinese Health and Nutrition Survey

The Chinese Health and Nutrition Survey (CHNS) is a large-scale, on-going longitudinal household survey in China, and it was first launched in 1989. The CHNS follows individuals randomly sampled from 228 communities, covering approximately 56 per cent of the Chinese population. A multi-stage cluster sample design is used to sample individuals within the provinces, and within these areas, neighbourhoods and households are randomly selected, following which all members of each household are interviewed.

The CHNS was designed to be representative of the nine provinces, but not designed to be nationally representative. However, the nine provinces covered vary substantially in terms of geography and economic development, and the CHNS has similar trends to other nationally representative surveys in China. A detailed description of the survey is available in Popkin, Du, Zhai, and Zhang (2010). The survey questions are very comprehensive, covering a household's economy, education, employment, consumptions, physical activities and a wide range of health conditions. The advantages of the CHNS for this analysis lie in its rich information about dietary choices, including nutrition and detailed information about daily physical activities, which are important observable behavioural variables associated with health conditions.

The eighth wave of the CHNS (i.e. CHNS 2009) takes detailed information about objective health status from the blood samples of the respondents, i.e. blood-based biomarkers. Individuals older than seven years visited a neighbourhood clinic to have trained physicians collect fasting blood samples. Following an overnight fast, blood was collected by venepuncture and tested immediately for glucose and haemoglobin A1c (HbA1c). Plasma and serum samples were then frozen and stored at −86°C for later laboratory analysis. All samples were analysed in a national central lab in Beijing along with strict quality controls. Individuals unable to attend the clinic had blood samples collected at home. This study concentrates on adult people aged over 19, because not all of the biomarkers for people aged below 19 are clinically comparable with those for the adult population. Pregnant women are dropped from the sample, because their biomarkers may not be comparable with those of non-pregnant women and men.

2.2. Variables

2.2.1. Health conditions

This study focuses on anthropometrics, blood pressure and the eight blood-based biomarkers. From measurements taken of the heights and weights of the respondents, the body mass index (BMI) was calculated, which is defined as an individual's weight divided by the square of his/her height and is expressed internationally in units of kg/m^2. Information regarding height and weight in the CHNS is based on actual measurements, and so they are less likely to be subject to reporting errors such as the under-reporting of weight (Gorber, Tremblay, Moher, & Gorber, 2007).[5] A high BMI measurement is a well-known indicator of being overweight and obese, but it is also associated with an increased risk of cardiovascular disease, stroke, diabetes and musculoskeletal disorders (Ng et al., 2014). Systolic blood pressure (systolic BP) and diastolic blood pressure (diastolic BP) used in this study are the mean values of the three measurements.[6] High blood pressure, i.e. hypertension, is also one of the most well-known causes of life-threatening complications such as heart attack and stroke (World Health Organization, 2013).

Besides these anthropometric and blood pressure measurements, the following markers were taken from the collected blood samples: triglycerides, haemoglobin A1c (HbA1c), glucose, uric acid, creatinine, high-density lipoprotein cholesterol (HDL cholesterol), low-density lipoprotein cholesterol (LDL cholesterol) and total cholesterol. Their related diseases are listed in Table 1.

Triglycerides are a type of fat (lipid) found in blood and measured in mg/dL, and they increase if more calories are consumed than a person burns on a regular basis. A high triglycerides level (sometimes called 'hypertriglyceridemia') is associated with a high risk of diabetes. Glycosylated haemoglobin (HbA1c) measures glucose metabolism in units called millimoles per litre of blood, $mmol/L$, and is often used to diagnose diabetes. A high level of HbA1c is indicative of the disease. Glucose, measured in mg/dL, is known as 'blood sugar', and a heavy meal, stress or lack of physical activity tend to raise its level. Insulin helps cope with this increased level of glucose; however, people with diabetes may need to administer synthetic insulin to deal with the issue. High glucose is an indication of diabetes, and not treating the problem could lead to neuropathy, heart disease, blindness and skin infections.

Uric acid in the blood, measured in mg/dL, is tested to help determine how well the body produces and removes uric acid. Uric acid is produced when the body breaks down foods containing purines, such as liver, anchovies, beer and wine. Most uric acid is dissolved in the blood and filtered through the kidneys, before being expelled in the urine. A high level of uric acid is associated with gout, while too little in the blood is an indication of possible liver or kidney diseases. Creatinine is a waste product in the blood, and it is removed therefrom by the kidney and then passed out of the body in urine. A high level of creatinine is associated with kidney diseases.

Cholesterol is a fatty substance known as a 'lipid' and is vital for the

[3] Yan et al. (2012) define the three levels of urbanicity from the urbanicity index derived by Jones-Smith and Popkin (2010).

[4] Specifically, Yan et al. (2012) estimate the odds ratio of the prevalences of being overweight, impaired-fasting glucose, diabetes, impaired A1c, high A1c, high triglycerides, high LDL cholesterol, low HDL cholesterol, high total cholesterol, hypertension, pre-hypertension and high CRP.

[5] In the survey, heights were measured (without shoes) by trained health workers to the nearest 0.1 cm with a portable SECA stadiometer (Seca North America East, Hanover, MD, USA), while weight was measured without shoes and in light clothing to the nearest 0.1 kg on a calibrated beam balance. Each of these measurements was taken by at least two health workers.

[6] Blood pressures were measured on the right arm, using mercury sphygmomanometers with appropriate cuff sizes. Measurements were collected in triplicate after a 10-min seated resting period.

Table 1
Objective health measurements.

Health measurements	Related health risks
Diabetes-related biomarkers	
BMI (kg/m^2)	Obesity, diabetes
Triglyceride (mg/dL)	Diabetes, hyperlipidaemia, liver ailment, etc.
HbA1c ($mmol/L$)	Impaired glucose control, diabetes
Glucose (mg/dL)	Impaired fasting glucose, diabetes
Cardiovascular and kidney-related biomarkers	
Systolic blood pressure ($mmHg$)	Hypertension
Diastolic blood pressure ($mmHg$)	Hypertension
Uric acid (mg/dL)	Gout, Kidney diseases
Creatinine (mg/dL)	Renal failure
Cholesterols	
HDL Cholesterol (mg/dL)	Hyperlipidaemia, low HDL
LDL Cholesterol (mg/dL)	Hyperlipidaemia, brain/cardiac infarction
Total cholesterol (mg/dL)	Hyperlipidaemia, diabetes, etc.

normal functioning of the body. Cholesterols are measured in mg/dL and are carried in the blood by proteins. When cholesterols and proteins are combined, they are called 'lipoprotein cholesterols', of which there are two main types: high-density lipoprotein (HDL) and low-density lipoprotein (LDL) cholesterols. HDL carries cholesterols away from the cells and back to the liver, and HDL cholesterol is often referred to as a 'good cholesterol', a high level of which is associated with better health conditions, whereas low HDL cholesterol is associated with a high risk of cardiovascular disease. LDL, on the other hand, carries cholesterols to the cells that need it; however, as too much cholesterol build-up can lead to artery disease, the LDL cholesterol is often known as a 'bad cholesterol'. Total cholesterols measure the total amount of cholesterols in the blood, including both LDL cholesterols and HDL cholesterols.

The shapes of the estimated density distribution of each health outcome are represented by histograms in Fig. 2. Most of the health outcomes exhibit distinctive features of non-normality. The non-parametric Kolmogorov-Smirnov test rejects the normality of all respective 11 health outcomes (all $p < 0.01$). We observe that systolic and diastolic blood pressure show some heaping points and that HbA1c, triglycerides and glucose show longer right tails.

2.3. Urbanicity index

Significant heterogeneity on the level of urban developments exists in China, not only between urban cities and rural villages, but also in communities within cities/villages. To take into account heterogeneity in urban developments, this study uses the urbanicity index, which quantifies the level of urban developments in a community and allows one to compare this level across these communities. Indexing the level of urban development provides a simple way to measure it (Galea & Vlahov, 2005), and achieving this aim through a scaler index allows for a finer level of investigation into the effect of urbanicity on health (Dahly & Adair, 2007; McDade & Adair, 2001). Cyril, Oldroyd, and Renzaho (2013) provide a systematic review of the urbanicity indices used in empirical studies in China and discuss their reliabilities and validities. According to Cyril et al. (2013), quantification of the urban environments has been implemented by Liu, Wu, Peng, and Fu (2003), Van de Poel et al. (2009), Monda et al. (2007) and Jones-Smith and Popkin (2010), who derived different indices by different methods; Cyril et al. (2013) rated their respective urbanicity indices from various perspectives and concluded that the urbanisation index derived by Jones-Smith and Popkin (2010) had the highest overall quality score.

This study therefore uses the urbanicity index defined by Jones-Smith and Popkin (2010), which allocates a maximum of 10 points to each of the 12 components of community-level characteristics.[7] The scale shows good internal consistency, and its reliability and validity are confirmed by a number of tests (Cyril et al., 2013). For a full explanation of this urbanisation index, see Jones-Smith and Popkin (2010). Fig. 3 illustrates the histogram of the urbanicity index and its kernel densities, estimated separately for rural and urban areas.[8] Fig. 3 shows the continuously distributed urbanicity index, ranging from 34.08 to 106.46, which clearly exhibits rural-urban overlapping and heterogeneities, both of which would not be captured by simple dichotomous rural-urban classification.

2.4. Control variables

This study uses two sets of control variables. The first is a set of exogenous control variables that affect health and health-related behaviours but are assumed not to be influenced by urban developments, while the other set is composed of observable behavioural variables that are likely to be influenced by urban developments.

As exogenous control variables, we use age and sex dummy variables. We also control for the Chinese unique legal household residency status (called *hukou*) to reflect the differences in available public policies. The *hukou* system was originally implemented in the 1950s, and under it each member of a household is legally bound to register his/her permanent place of residence and type of residency, which is either rural or urban, based on the mother's registration status at birth (Chan & Zhang, 1999). The system connects to the government's social programmes, which range from education and healthcare to retirement pension. Most importantly, the *hukou* system actively restricts where a person is allowed to live, especially if one is born into a rural *hukou*. This limitation was designed to prevent mass rural-urban labour migration, which used to be essential in the centrally-planned economy (Liu, 2005).

The selection of the behavioural variables in this study is supported by evidence in the empirical literature on urban health in China, as well as the epidemiological and public health literature on NCDs (e.g. Aizawa, 2018a; French, Story, & Jeffery, 2001; Kim, 2016; Konteh, 2009; Moore et al., 2003; Popkin, Kim, Rusev, Du, & Zizza, 2006; Swinburn, Caterson, Seidell, & James, 2004; Tucker & Kano, 1992), and those used in this study are family size, marital status, occupation types, household income, individual education attainment, consumption patterns and daily physical activities.

First, for occupation types, we use the following dummy variables: professional worker, farmer, self-employed worker, permanent worker, contractor and temporary worker. Second, for the affluence of a household, we use per capita household net income (yuan), taken from a wide range of sources, including salaries, household business, income in kind and subsidies. For educational background, we include years of education. Consumption patterns consist of energy intake ($kcal$), fat intake (g), protein intake (g) and carbohydrate intake (g), which are calculated by the National Institute for Nutrition and Health on the basis of the average of three days' consumption of meals and snacks (Popkin et al., 2010).

Lastly, we take account of the strength of daily physical activities. It is important to include various types of physical activities commonly observed in daily life. To make the strength of different daily physical activities comparable, they are quantified as the time spent weighted by the strength of activities and measured by the metabolic equivalent (MET). MET is a physiological measure that expresses energy expended due to physical activities. A unit of MET is defined as the ratio of the metabolic rate during a specific physical activity to a reference

[7] These are: population density, economic activity, traditional markets, modernity, transportation infrastructure, sanitation, communications, housing, education, diversity, health infrastructure and social services.

[8] The urban-rural definition in Fig. 3 is based on the definition set by the CHNS.

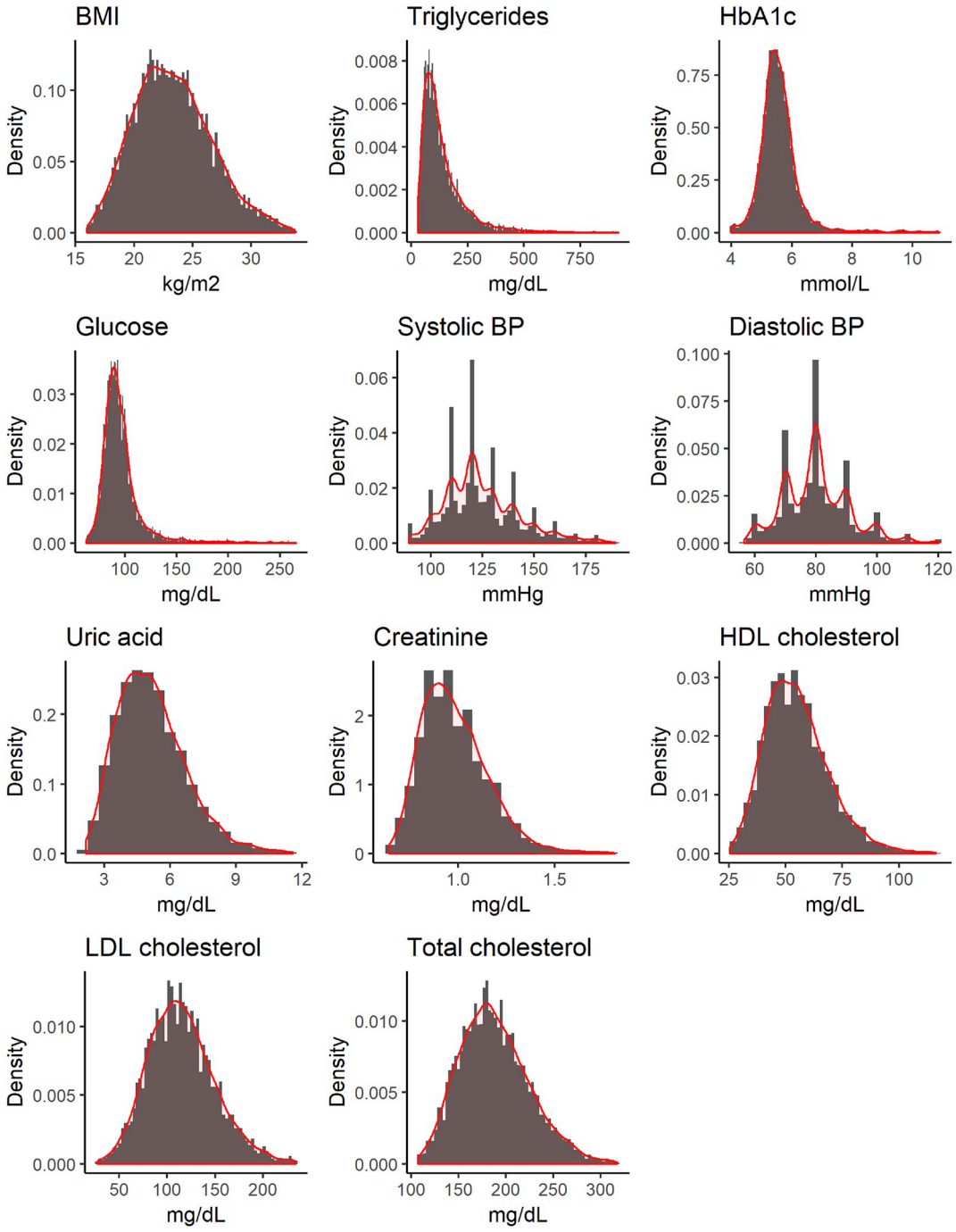

Fig. 2. Health outcome distributions.

metabolic rate at resting (Sallis et al., 1985). Following Ng, Norton, and Popkin (2009), in order to quantify the strength of physical activities and to make them comparable across different types of activities, time spent on each activity per week is multiplied by specific MET values on the compendium of physical activities (Ainsworth et al., 1993, 2011).

METs for activities at home are calculated according to the time spent preparing food, buying food, doing the laundry and childcare: 2.3 METs for buying food for a household, 2.25 METs for preparing food or cooking for a household, 2.15 METs for doing the laundry, 2.5 METs for cleaning the house and 2.75 METs for childcare. Next, for leisure activities, the following METs per hour are assigned: 4.5 for martial arts, 7.5 for track and field (running etc.) or swimming, 5 for gymnastics, dancing or aerobics, 6 for playing basketball, volleyball, football, tennis or badminton and 5 for other sports such as ping-pong and Tai Chi. For transportation activities, 1.5 METs are allocated for taking a motorised vehicle, 4 METs for cycling and 3 METs for walking. Occupational activities are measured by the respondent's occupation and the average number of hours spent working per week in the previous year for up to two market sector jobs as well as hours worked at home. For farming, fishing, hunting, working in a garden or orchard and working with livestock, 6 METs are assigned. For skilled/non-skilled workers,[9] service workers[10], drivers, ordinary soldiers, policemen, athletes, actors,

[9] These are foremen, group leaders, craftsmen, ordinary labourers and loggers.

[10] These are housekeepers, cooks, waiters, doorkeepers, hairdressers, counter salespersons, laundresses and childcare workers.

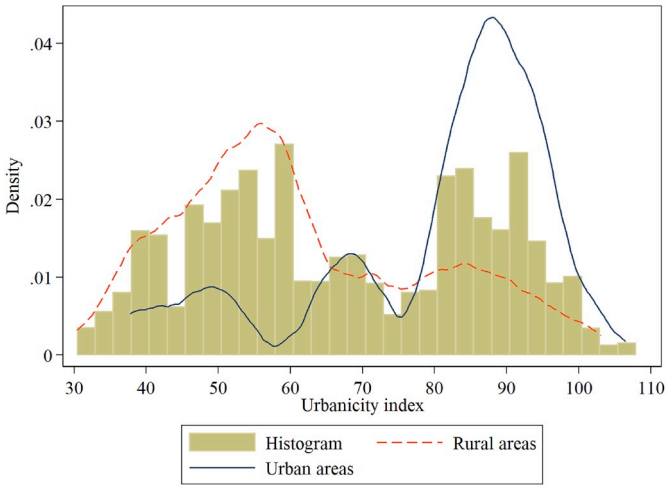

Fig. 3. Urbanicity index densities.

musicians and students, 4 METs per hour are assigned. Two METs are assigned to senior/junior professionals[11] and technical workers,[12] administrators, executives, managers and office staff.[13] We aggregated the METs from all of the activities mentioned above.

2.5. Sample selection

The original sample size was 11,887. First, 1,957 observations of people aged below 20 years are dropped, and 79 observations of pregnant women are dropped. Next, by dropping incomplete observations with missing control variable values, the sample size becomes 9,205.[14] We then drop samples with possible outliers.[15] To make the most of the available information, our final sample size for the analysis varies across health outcomes, ranging from 7,631 observations for systolic blood pressure, to 8,625 observations for BMI. We do not observe any statistically significant differences in the means of health outcomes between the dropped samples and the included samples (Table A.2 in Appendix), thus implying that dropping observations with missing information on control variables is not likely to produce selection bias. The descriptive statistics of the health variables and control variables are shown in Table 2.

3. Methodology

3.1. Counterfactual health distribution

This study uses the following notations. H denotes a continuous health variable with support $\mathscr{H} \subset \mathbb{R}$, and $X = \{X_1, ..., X_K\}$ is a vector of health-related behavioural variables with support $\mathscr{X} \subset \mathbb{R}^K$. $Z = \{Z_1, ..., Z_P\}$ is a vector of exogenous determinants of health and health-related behaviours with support $\mathscr{Z} \subset \mathbb{R}^P$. U is the scalar urbanicity index with support $\mathscr{U} \subset \mathbb{R}$. H, X, Z and U have distribution F_H, F_X, F_Z and F_U, respectively. In this study, X is assumed to be influenced by

Table 2
Descriptive statistics.

	count	mean	sd	min	max
BMI	8625	23.33	3.32	16.02	33.73
Triglyceride (mg/dL)	7856	143.32	107.59	31.89	919.40
HbA1c (mmol/L)	7821	5.60	0.75	4.00	10.90
Glucose (mg/dL)	7854	96.54	21.79	62.69	265.84
Systolic BP (mmHg)	7631	124.68	18.03	89.33	189.33
Diastolic BP (mmHg)	7651	80.33	10.78	56.67	120.00
Uric Acid (mg/dL)	7849	5.15	1.56	2.17	11.60
Creatinine (mg/dL)	7860	0.98	0.17	0.64	1.81
HDL cholesterol (mg/dL)	7858	55.01	14.01	25.52	116.40
LDL cholesterol (mg/dL)	7852	115.17	34.49	27.84	233.95
Total cholesterol (mg/dL)	7855	188.07	36.79	107.50	317.09
Urbanicity index	9205	67.71	19.46	30.42	106.46
Age 30-39	9205	0.16	0.36	0.00	1.00
Age 40-49	9205	0.23	0.42	0.00	1.00
Age 50-59	9205	0.24	0.43	0.00	1.00
Age over 60	9205	0.28	0.45	0.00	1.00
Male	9205	0.48	0.50	0.00	1.00
Urban registration	9205	0.42	0.49	0.00	1.00
Married	9205	0.84	0.37	0.00	1.00
Family size	9205	3.12	1.48	1.00	10.00
Professional worker	9205	0.07	0.26	0.00	1.00
Farmer	9205	0.27	0.44	0.00	1.00
Self-employed/business owner	9205	0.35	0.48	0.00	1.00
Permanent worker	9205	0.12	0.33	0.00	1.00
Contractor	9205	0.05	0.22	0.00	1.00
Temporary worker	9205	0.06	0.23	0.00	1.00
Family size	9205	3.12	1.48	1.00	10.00
Education years	9205	7.29	4.71	0.00	16.00
Household income (divided by 1,000)	9205	10.27	9.61	0.00	77.92
Energy intake (kcal)	9205	2105.12	605.32	587.55	4188.97
Fat intake (g)	9205	73.38	33.87	8.23	209.66
Protein intake (g)	9205	65.29	21.55	16.91	143.36
Carbohydrate intake (g)	9205	290.90	97.67	76.00	615.03
METs	9205	118.89	119.01	0.00	558.51

Age, sex, registarion, marital status, and occupation types are binary variables.

(Z, U), but Z is not influenced by U. The lower cases with subscript i, such as h_i and x_i, denote the individual i's realised values.

Using the law of iterated probabilities, the distribution of H, is expressed by

$$F_H(h) = \int_{\mathscr{Z}} \int_{\mathscr{U}} \int_{\mathscr{X}} F_{H|X,Z,U}(h) dF_{X|Z,U}(x) dF_{U|Z}(u) dF_Z(z), \quad (1)$$

where $F_{H|X,Z,U}$ is a conditional distribution of H, given X, Z and U, $F_{X|Z,U}$ is a conditional distribution of X, given Z and U, and $F_{U|Z}$ is a conditional distribution of U, given Z. Let $v(F_H)$ denote the distributional statistics of H, where v: $F_H \mapsto \mathbb{R}$ is a functional from the space of a one-dimensional distribution function to the real line. For example, for the τ_{th} quantile of health, v: $F_H \mapsto F_H^{-1}(\tau)$, $\tau \in [0,1]$ and for the mean v: $F_H \mapsto \int_{\mathscr{H}} h dF_H(h)$. We would now like to infer the effect of urban developments on health distribution, by comparing the distributional statistics of the observed health outcomes with those of the counterfactual health outcomes in a situation where every individual i is exposed to the lowest level of urban development. In other words, in this counterfactual situation, the urbanicity index is fixed at its empirical minimum level. This counterfactual health distribution can be estimated by manipulating the distribution of U in such a way that makes $\forall\, i$, $u_i = \underline{u}$, where $\underline{u}$ denotes an observed minimum value of U. The distribution of X can change in accordance with the manipulation of the distribution of U. For example, household income and consumption patterns are likely to change when the level of urban developments changes, and these changes are then likely to affect health distribution. Taking into account this mechanism, through which a change in the distribution of the urbanicity index affects the distribution of behavioural variables, we introduce a counterfactual distribution of health in which U is fixed to $\underline{u}$ and X follows counterfactual

[11] Namely doctors, professors, lawyers, architects and engineers.
[12] Midwives, nurses, teachers, editors and photographers.
[13] These are secretaries and office helpers.
[14] Twelve observations have missing information on occupation type, 219 observations have missing information on household income, 342 observations have missing information on consumption patterns and 73 observations have missing information on physical activities.
[15] Some observations show extremely high or low biomarkers, many of which seem biologically impossible. Observations sitting in the top 0.1% and bottom 0.1% of each health outcome are dropped from the sample as possible outliers. In this process, around 80 observations are dropped. Cut-off values are provided in Table A.1 in Appendix.

distribution, conditional on $U = \underline{u}$ as follows:

$$F_H^{CF}(h) = \int_{\mathscr{Z}} \int_{\mathscr{X}} F_{H|X,Z,U=\underline{u}}(h) dF^{CF}_{X|Z,U=\underline{u}}(x) dF_Z(z), \tag{2}$$

where $F^{CF}_{X|Z,U=\underline{u}}(x)$ is a counterfactual conditional distribution of X, given Z and $U = \underline{u}$.

The effect of manipulating U on the distributional statistics of health is given by

$$\Delta v(F_H) = v(F_H) - v(F_H^{CF}). \tag{3}$$

We call $\Delta v(F_H)$ the 'total effect' of urbanicity, which captures the overall effect of urban developments.

It is also interesting to know how much part of the overall effect is driven by the channel by which manipulating the distribution of U influences the distribution of X. We call this effect the 'behavioural effect', whist we refer to the remaining part that cannot be explained by a change in X the 'non-behavioural effect', which can also be interpreted as the unobserved effect in the sense that the effect of the manipulation is not attributable to change in the distribution of observed behaviours. We estimate these two effects by introducing another type of counterfactual distribution in which U is fixed to $\underline{u}$, and yet X follows the original observed distribution:

$$F_H^{CF2}(h) = \int_{\mathscr{Z}} \int_{\mathscr{X}} F_{H|X,Z,U=\underline{u}}(h) dF_{X|Z}(x) dF_Z(z). \tag{4}$$

This second type of counterfactual health distribution does not take into account the pathway through which the manipulation of U affects H via X. Using this second type of counterfactual health distribution, we can distinguish the pathways by which U influences H by decomposing the equation (3) as follows:

$$\Delta v(F_H) = \underbrace{v(F_H) - v(F_H^{CF2})}_{Non-behavioural\ effect} + \underbrace{v(F_H^{CF2}) - v(F_H^{CF})}_{Behavioural\ effect}. \tag{5}$$

In equation (5), the behavioural effect captures the part of the total effect that is due to the change in the distribution of health-related behaviours triggered by the manipulation of U, because the behavioural effect is based on the differences between the distribution of observed behavioural variables $F_{X|Z,U}$ and those of counterfactual behavioural variables $F_{X|Z,U=\underline{u}}$. The non-behavioural effect is the remaining part that is not attributable to the observed behavioural change.

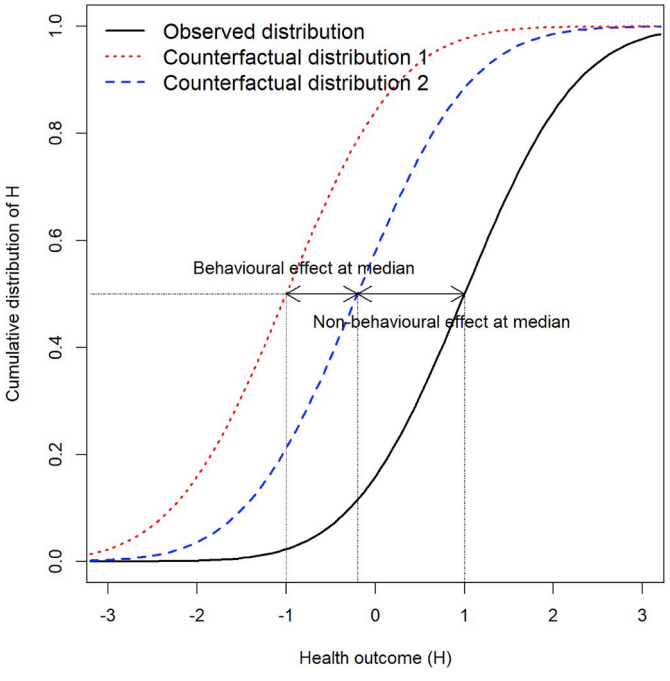

Fig. 4. Decomposition of the total effect.

Fig. 4 illustrates the idea of decomposing the total effect, whereby a total effect at median is decomposed into non-behavioural and behavioural effects. First, the total effect is measured by the horizontal difference between the observed distribution (F_H) and the first-type counterfactual distribution (F_H^{CF}). Next, the non-behavioural effect is measured by the horizontal difference between the observed distribution (F_H) and the second-type counterfactual distribution (F_H^{CF2}). Lastly, the behavioural effect is measured by the horizontal difference between the first-type counterfactual distribution (F_H^{CF}) and the second-type counterfactual distribution (F_H^{CF2}). The effects' sizes can vary across the distribution, and we investigate the heterogeneous size of each one across the health distribution.

3.2. Estimation

Estimating total, behavioural and non-behavioural effects requires estimating the counterfactual distributions of X and H. We first estimate the counterfactual distribution of X, and then we estimate the counterfactual distribution of H.

3.2.1. Counterfactual distribution of behavioural variables

We first estimate the conditional mean function of $X_k \in X$, $(k = 1, ..., K)$, given Z and U. We employ the Box-Cox transformation for X_k to deal with the non-linearity of its conditional mean function (Box & Cox, 1964). The Box-Cox transformation produces a conditional distribution of the transformed outcome variable more closely resembling normality and exhibiting less heteroskedasticity. The transformed outcome $x_{ki}^{(\theta_k^X)}$ is given by

$$x_{ki}^{(\theta_k^X)} = \begin{cases} \frac{x_{ki}^{\theta_k^X} - 1}{\theta_k^X}, & if\ \theta_k^X \neq 0 \\ ln(x_{ki}), & if\ \theta_k^X = 0, \end{cases} \tag{6}$$

where θ_k^X is a Box-Cox transformation parameter for X_k, and it can be estimated by applying the maximum likelihood method. The transformation embeds several popular functional forms, including levels ($\theta_k^X = 1$), square root ($\theta_k^X = \frac{1}{2}$) and logs ($\theta_k^X = 0$) as special cases. Next, we regress the transformed outcome on Z, U and their interaction terms.

$$x_{ki}^{(\theta_k^X)} = \beta_0^k + \sum_{p=1}^{P} \beta_{Z_p}^k z_{pi} + \beta_U^k u_i + \beta_{U^2}^k u_i^2 + \beta_{U^3}^k u_i^3 + \sum_{p=1}^{P} (\beta_{UZ_p}^k) u_i * z_{pi} + \varepsilon_i^k, \tag{7}$$

where ε_i^k is an unobservable factor. We specify the conditional mean function as a cubic function of U to take account of a possible non-linear relationship between X and U. Including two-way interaction terms between U and Z captures heterogeneous associations between X and U across the values of Z. ε_i^k is assumed to be composed of both individual-specific unobservable characteristics and idiosyncratic errors ($\varepsilon_i^k = \alpha_i^k + \kappa_i^k$).

Then, for each X_k, we estimate the coefficients, $\widehat{\beta_0^k}$, $\widehat{\beta_{Z_p}^k}$, $\widehat{\beta_U^k}$, $\widehat{\beta_{U^2}^k}$, $\widehat{\beta_{U^3}^k}$ and $\widehat{\beta_{UZ_p}^k}$, and residuals, $\widehat{\varepsilon_i^k}$. Note that due to the possible endogeneity of U, the estimated coefficients $\widehat{\beta_U^k}$, $\widehat{\beta_{U^2}^k}$ and $\widehat{\beta_{U^3}^k}$ do not necessarily capture the causal impact of U. Therefore, they show only the correlation between X_k and U.

Next, we predict the counterfactual value of $X_k^{(\theta_k^X)}$ for each individual i, which we denote as $\widetilde{x_{ki}}^{(\theta_k^X)}$. Counterfactual distribution, in the situation where $u_i = \underline{u}$, can be obtained under the assumptions that (1) manipulating the distribution of U changes neither individual unobservable characteristics nor idiosyncratic errors (stability assumption) and (2) $\underline{u}$ is in the support of U (support assumption). Since, in this study, $\underline{u}$ is an observed minimal value of U, the support assumption is trivially satisfied. The stability assumption requires that any change in the distribution of U does not change the estimated coefficients and residuals. These assumptions allow us to interpolate counterfactual

distributions from the data. $\widetilde{x}_{ki}^{(\theta_k^X)}$ is predicted as follows:

$$\widetilde{x}_{ki}^{\left(\theta_k^X\right)} = \widehat{\beta_0^k} + \sum_{p=1}^{P} \widehat{\beta_{Zp}^k} z_{pi} + \widehat{\beta_U^k} \underline{u} + \widehat{\beta_{U^2}^k} \underline{u}^2 + \widehat{\beta_{U^3}^k} \underline{u}^3 + \sum_{p=1}^{P} \left(\widehat{\beta_{UZp}^k}\right) \underline{u}^* z_{pi} + \widehat{\varepsilon_i^k}. \tag{8}$$

By adding $\widehat{\varepsilon_i^k}$, we reflect the unobservable individual-specific characteristics to the predicted value of $X_k^{(\theta_k^X)}$. The important point to note is that the predicted counterfactual distribution reflects the original correlation between U and α, which implies that the difference between $x_{ki}^{(\theta_k^X)}$ and $\widetilde{x}_{ki}^{(\theta_k^X)}$ does not necessarily capture the causal effect of urbanicity. The difference reflects the net effect of urbanicity on $X_k^{(\theta_k^X)}$ that reflects individual-specific characteristics. Repeating this estimation for every $k = 1, \ldots, K$, we can obtain the counterfactual distribution of $X^{(\theta^X)}$.

When X_k is a binary variable, we estimate the coefficients by employing the logit model and predicting the counterfactual probability for each individual. The binary-response model ensures that the predicted probabilities range between 0 and 1.[16]

3.2.2. Counterfactual distribution of a health variable

Next, we estimate the conditional distribution function of H and predict its counterfactual distribution. The conditional mean function of H can be estimated in a fashion similar to how we estimated the conditional mean function of X. Again, we first transform H with the Box-Cox transformation, and so the transformed health outcome, $h_i^{(\theta^H)}$, is given by

$$h_i^{(\theta^H)} = \begin{cases} \frac{h_i^{\theta^H} - 1}{\theta^H}, & \text{if } \theta^H \neq 0 \\ ln(h_i), & \text{if } \theta^H = 0, \end{cases} \tag{9}$$

where θ^H is a Box-Cox transformation parameter for H. Then we regress the transformed outcome on the transformed X, Z and U and their interaction terms.

$$h_i^{(\theta^H)} = \gamma_0 + \sum_{k=1}^{K} \gamma_{X_k} x_{ki}^{(\theta_k^X)} + \sum_{p=1}^{P} \gamma_{Z_p} z_{pi} + \gamma_U u_i + \gamma_{U^2} u_i^2 + \gamma_{U^3} u_i^3 + \sum_{k=1}^{K} (\gamma_{UX_k}) u_i^* x_{ki}^{(\theta_k^X)} + \sum_{p=1}^{P} \left(\gamma_{UZ_p}\right) u_i^* z_{pi} + \eta_i, \tag{10}$$

where η_i is an error term and it is assumed to be composed of both the unobservable characteristics of individuals, such as genetic characteristics and idiosyncratic errors. Including the polynomial function of U can capture a possible non-linear relationship between H and U.

Using the estimated coefficients and residuals, we predict counterfactual health distribution in a situation where every individual i is exposed to the lowest observed level of urban development. Again, under the stability assumption – requiring that manipulating the distribution of U does not influence the estimated parameters[17] – and the common support assumption, we predict the counterfactual transformed health distribution by substituting $u_i = \underline{u}$ and $x_{ki}^{(\theta_k^X)} = \widetilde{x}_{ki}^{(\theta_k^X)}$ into the estimated model, thereby obtaining

$$\widetilde{h}_i^{(\theta^H),CF} = \widehat{\gamma_0} + \sum_{k=1}^{K} \widehat{\gamma_{X_k}} \widetilde{x}_{ki}^{(\theta_k^X)} + \sum_{p=1}^{P} \widehat{\gamma_{Z_p}} z_{pi} + \widehat{\gamma_U} \underline{u} + \widehat{\gamma_{U^2}} \underline{u}^2 + \widehat{\gamma_{U^3}} \underline{u}^3 + \sum_{k=1}^{K} (\widehat{\gamma_{UX_k}}) \underline{u}^* \widetilde{x}_{ki}^{(\theta_k^X)} + \sum_{p=1}^{P} \left(\widehat{\gamma_{UZ_p}}\right) \underline{u}^* z_{pi} + \widehat{\eta_i}. \tag{11}$$

$\tilde{h}_i^{(\theta^H),CF}$ in equation (11) is a predicted transformed counterfactual health outcome for individual i. By including the residual $\hat{\eta}_i$, the predicted counterfactual health distribution takes account of individual unobservable characteristics. Using the predicted counterfactual distribution of behavioural variables, i.e. $\widetilde{x}_{ki}^{(\theta_k^X)}$, the predicted counterfactual health distribution takes into account the behavioural changes through which urban developments influence health.

Next, we predict second-type counterfactual health distribution, which does not take into account behavioural changes due to urban developments. Second-type counterfactual health distribution is obtained by

$$\widetilde{h}_i^{(\theta^H),CF2} = \widehat{\gamma_0} + \sum_{k=1}^{K} \widehat{\gamma_{X_k}} x_{ki}^{(\theta_k^X)} + \sum_{p=1}^{P} \widehat{\gamma_{Z_p}} z_{pi} + \widehat{\gamma_U} \underline{u} + \widehat{\gamma_{U^2}} \underline{u}^2 + \widehat{\gamma_{U^3}} \underline{u}^3 + \sum_{k=1}^{K} (\widehat{\gamma_{UX_k}}) \underline{u}^* x_{ki}^{(\theta_k^X)} + \sum_{p=1}^{P} \left(\widehat{\gamma_{UZ_p}}\right) \underline{u}^* z_{pi} + \widehat{\eta_i}. \tag{12}$$

Note that in predicting this second-type counterfactual health distribution in equation (12), we do not use the predicted counterfactual distribution of X but use the original transformed value of X, i.e. $X_k^{(\theta^X)}$. Hence, in contrast to $\tilde{h}_i^{(\theta^H),CF}$ in equation (11), $\tilde{h}_i^{(\theta^H),CF2}$ in equation (12) does not consider the observed mechanism by which urbanicity influences health through its effect on X. This difference allows us to estimate the part of the total effect that can be attributed to the changes in behavioural variables due to urban developments. Lastly, by re-transforming $\tilde{h}_i^{(\theta^H),CF}$ and $\tilde{h}_i^{(\theta^H),CF2}$ to their original scale, we can obtain predicted counterfactual health outcomes for individual i.[18] Using these predicted health outcomes for each individual, we then estimate the total, non-behavioural and behavioural effects via equation (5).

4. Results

4.1. Counterfactual behavioural variables

Table 3 compares the descriptive statistics of the observed X variables with those of the counterfactual X variables. We re-transform the predicted values, $\tilde{X}^{(\theta^X)}$ to the original scale so that we can compare the distributions.[19] The regression results from which the counterfactual distributions are estimated are provided in Table A.4 in Appendix. Looking at the counterfactual behavioural variables, we find significant higher values of average energy intake ($p < 0.05$), carbohydrate intake ($p < 0.01$) and MET values ($p < 0.01$), which means that if every individual were exposed to the lowest level of urban development, he/she, on average, would increase their energy and carbohydrate intake and become more physically active. On the other hand, significant declines in household income, education attainment, fat intake and protein intake are found (all $p < 0.01$). Furthermore, we observe reductions in the probabilities of being a professional worker, a

[16] We also estimate the counterfactual probabilities with a linear probability model (not shown), albeit the final results are very similar to the ones reported in this paper.

[17] This assumption could be violated if urban developments are expected to change the distribution of individual unobservable characteristics. For example, if hormone-disrupting chemicals produced through the process of urban developments substantially change individual genetic characteristics, the predicted counterfactual distribution could suffer prediction errors.

[18] Specifically, re-transformation can be achieved by $\widetilde{h}_i = \left(\widetilde{h}_i^{(\theta^H)} \theta^H + 1\right)^{\frac{1}{\theta^H}}$, if $\theta^H \neq 0$, and $\widetilde{h}_i = exp\left(\widetilde{h}_i^{(\theta^H)}\right)$, if $\theta^H = 0$.

[19] Specifically, for all $k \in K$, re-transformation can be achieved by $\widetilde{x}_{ki} = \left(\widetilde{x}_{ki}^{\left(\theta_k^X\right)} \theta_k^X + 1\right)^{\frac{1}{\theta_k^X}}$, if $\theta_k^X \neq 0$, and $\widetilde{x}_{ki} = exp\left(\widetilde{x}_{ki}^{\left(\theta_k^X\right)}\right)$, if $\theta_k^X = 0$.

Table 3
Descriptive statistics for the observed and counterfactual behavioural variables.

	mean	sd	skewness	kurtosis	p50
Observed					
Married	0.84	0.37	−1.84	4.40	1.00
Professional worker	0.07	0.26	3.26	11.66	0.00
Farmer	0.27	0.44	1.05	2.09	0.00
Self-employed/business owner	0.35	0.48	0.65	1.42	0.00
Permanent worker	0.12	0.33	2.32	6.38	0.00
Contractor	0.05	0.22	4.05	17.43	0.00
Temporary worker	0.06	0.23	3.85	15.80	0.00
Family size	3.12	1.48	1.04	4.06	3.00
Household income (divided by 1,000)	10.27	9.61	2.25	10.45	7.50
Education years	7.29	4.71	−0.31	2.19	9.00
Energy intake (kcal)	2105.12	605.32	0.41	3.00	2058.21
Fat intake (g)	73.38	33.87	0.76	3.67	68.52
Protein intake (g)	65.29	21.55	0.65	3.36	62.36
Carbohydrate intake (g)	290.90	97.67	0.53	3.07	278.80
METs	118.89	119.01	1.00	3.34	86.00
Observations	9205				
Counterfactual					
Married	0.86	0.12	−0.53	1.46	0.92
Professional worker	0.05	0.05	1.20	3.22	0.03
Farmer	0.55	0.28	−0.28	1.53	0.64
Self-employed/business owner	0.55	0.29	−0.26	1.48	0.62
Permanent worker	0.06	0.07	1.81	6.28	0.03
Contractor	0.04	0.05	2.26	9.39	0.02
Temporary worker	0.05	0.04	1.36	4.20	0.04
Family size	3.09	1.47	1.02	3.97	2.81
Household income (divided by 1,000)	7.47	7.32	2.85	15.97	5.44
Education years	6.04	3.65	0.05	2.22	6.28
Energy intake (kcal)	2240.27	624.67	0.39	3.03	2195.21
Fat intake (g)	68.13	32.66	0.80	3.78	63.55
Protein intake (g)	63.02	21.10	0.66	3.39	60.20
Carbohydrate intake (g)	338.89	103.11	0.46	3.05	329.32
METs	127.77	131.56	1.15	3.77	83.70
Observations	9205				

permanent worker, a contractor or a temporary worker (all $p < 0.01$). On the other hand, increases in the proportions of farmers and self-employed workers are also observed (both $p < 0.01$). We find no significant difference in family size at the 5 per cent level, while the non-parametric Kolmogorov-Smirnov tests reject the equalities of the observed and counterfactual distributions for all respective behavioural variables at the 1 per cent level.

4.2. Total, non-behavioural, and behavioural effects

Next, Fig. 5 summarises visually the results by plotting the estimated total, non-behavioural and behavioural effects from the 5th to the 95th percentiles of the respective health variables. The regression results from which the counterfactual health distributions are estimated are provided in Table A.5 in Appendix. In the following subsections, we look closely at the three effects for each health outcome.

4.2.1. Diabetes-related biomarkers: BMI, triglycerides, HbA1c and glucose

Columns 1–6 in Table 4 show the quantiles of the observed and counterfactual distributions of BMI, triglycerides, HbA1c and glucose. Columns 7–12 in Table 4 show estimates of the total, non-behavioural and behavioural effects. Associated standard errors are obtained by bootstrap with 500 repetitions. The p-values come from testing the hypothesis that respective effects are zero. First, for BMI, we find a significant negative total effect across the distribution except at the left tail of the distribution ($p < 0.10$). The negative total effect means that we would find an increase in BMI if the urban development level were fixed at its observed minimum level. We find that a negative non-behavioural effect is offset by a positive behavioural effect, the latter of which suggests that changes in observed behavioural variables, due to urban developments, contribute to an increase in BMI, which in turn implies that the significant increases in household income, educational attainment, fat intake, significant decrease in physical activities and change in occupational choice, in the process of urban development, contribute to the rise in body mass. Moreover, a larger behavioural effect is observed at the higher percentile of the distribution (Fig. 5), which suggests that the behavioural changes are larger among overweight/obese people and/or that overweight/obese people are more susceptible to behavioural changes.

For triglycerides, we observe a significant positive total effect for most parts of the distribution ($p < 0.05$). We find a substantially larger total effect on the higher percentiles of the distribution (Fig. 5), and the largest effect is observed at the right tail thereof. This substantial heterogeneity across the distribution implies that urban developments have a stronger adverse effect among those people sitting at the higher percentiles of the distribution, and these urban developments change the shape of the distribution. The total effect is explained by both non-behavioural and behavioural effects, both of which show heterogeneity in their sizes across the distribution, i.e. larger effects on the right tail of the distribution. However, we observe large standard errors for the non-behavioural effect, thereby rendering the non-behavioural effect statistically insignificant.

For HbA1c, a significant negative total effect is observed across the distribution ($p < 0.05$), most of which is due to the non-behavioural effect. We do not observe a significant behavioural effect, which means that the change in the distribution of HbA1c is not explained by the observed change in behavioural variables due to urban developments. For glucose, we find a significant total effect across the distribution ($p < 0.01$), and its largest effect is observed on the right tail thereof (Fig. 5). Most parts of the total effect are attributable to the non-behavioural effect, and we do not observe a significant behavioural effect. These results, for the cases of HbA1c and glucose, suggest that the changes in their distributions are triggered by unobservable behavioural changes and/or changes in individual non-behavioural characteristics.

4.2.2. Cardiovascular and kidney-related biomarkers: systolic BP, diastolic BP, uric acid and creatinine

Table 5 shows the effects on the cardiovascular and kidney-related biomarkers. First, for systolic blood pressure, we find a significant positive total effect across the distribution ($p < 0.05$). This larger effect is observed on the higher percentiles. The size of the behavioural effect is relatively constant across the distribution, thereby suggesting that behavioural changes shift the location of the distribution of systolic blood pressure without changing its shape. However, at the same time, the behavioural effect exhibits a large standard error across the distribution. The non-behavioural effect shows a larger contribution to the total effect across the distribution (Fig. 5), meaning that much of the change in the distribution of systolic blood pressure is not explained by the change in the distribution of behavioural variables.

For diastolic blood pressure, the total effect is positive and significant at most percentiles of the distribution ($p < 0.05$), and its large effect is observed on the higher percentiles thereof. Consistent with systolic blood pressure, the size of the total effect shows some bumpy spikes, and this is possibly because of some heaping points observed in the original distribution (Table 2). Similar to the results for systolic blood pressure, the size of the behavioural effect is relatively constant across the distribution and exhibits a large standard error.

For uric acid, we observe a positive total effect for all parts of the distribution, the size of which is larger on the higher percentiles (Fig. 5), though the positive effect exhibits a large standard error relative to the effect size, which makes the total effect statistically insignificant. We find that a negative non-behavioural effect is offset by a positive behavioural effect, both of which are statistically significant at all estimated percentiles of the distribution ($p < 0.01$). The absolute size

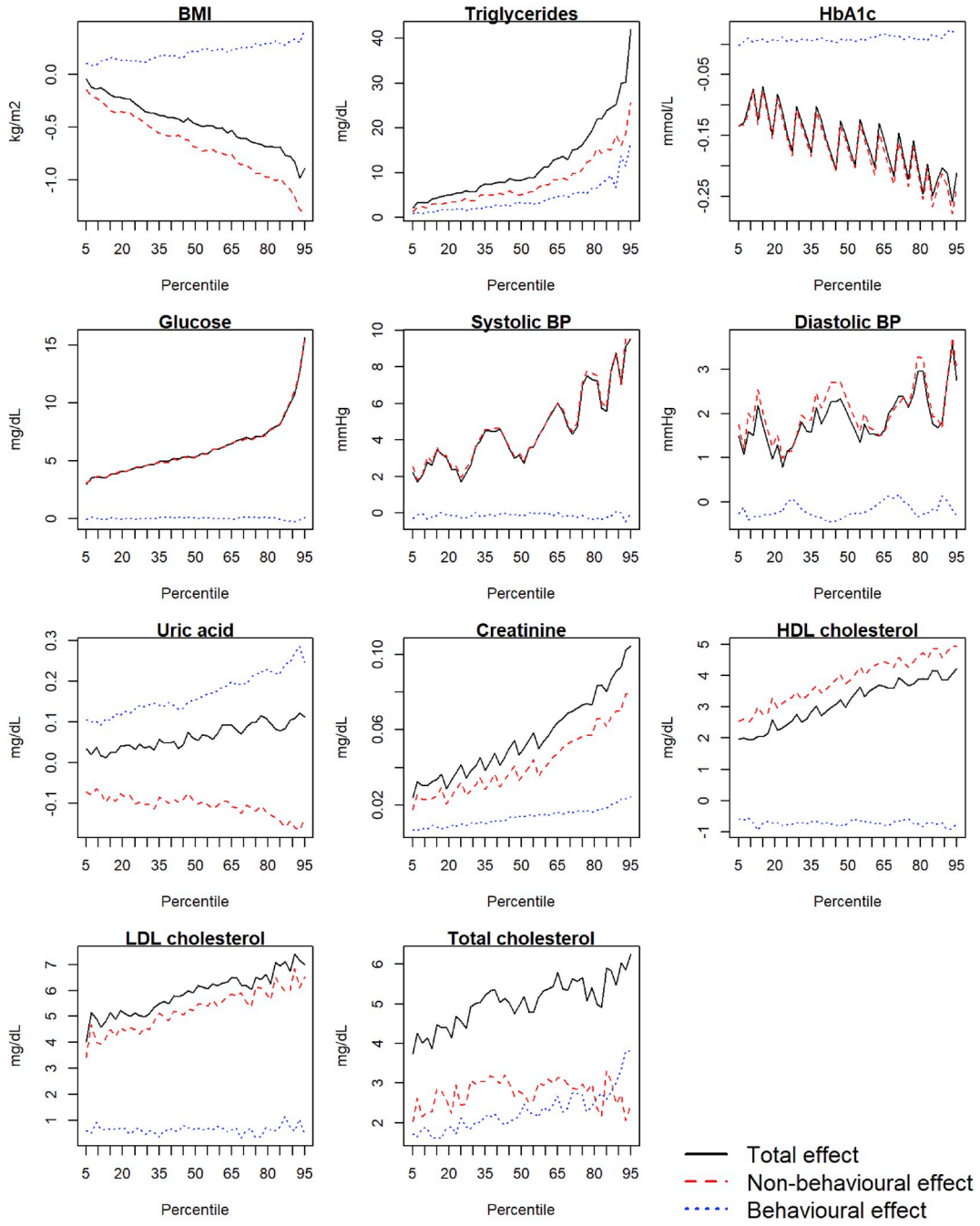

Fig. 5. Total, non-behavioural and behavioural effects.

of the behavioural effect is larger than that of the non-behavioural effect, thereby producing the positive total effect. The behavioural effect is larger on the higher percentiles of the distribution, which implies that behavioural changes are larger among people with a high uric acid level and/or that people sitting at the higher percentile of the distribution are more susceptible to behavioural changes.

For creatinine, we observe a positive and significant total effect on all estimated percentiles ($p < 0.01$), with the larger total effect at the higher percentile points of the distribution. The largest total effect is observed at the right tail of the distribution. A larger proportion of the total effect is attributable to the significant positive non-behavioural effect ($p < 0.01$). Although both non-behavioural and behavioural effects show larger effects on the higher percentiles of the distribution, the degree of heterogeneity in the behavioural effect across the distribution is much smaller than that of the total and non-behavioural effects. Hence, the change in the distribution of behavioural variables explains the total effect less so for the higher percentile of the distribution.

4.2.3. *Cholesterols: HDL cholesterol, LDL cholesterol and total cholesterol*

Table 6 shows the effects on the cholesterols. First, for HDL cholesterol, we find a significant positive total effect across the distribution ($p < 0.01$), and its size becomes larger on the higher percentiles of the distribution. We find a positive non-behavioural effect and a negative behavioural effect. The absolute size of the non-behavioural effect is far larger than that of the behavioural effect, thereby producing the positive total effect. Moreover, while the size of the non-behavioural effect exhibits heterogeneity across the distribution, that of the behavioural effect is relatively constant. The negative and constant behavioural effects mean that the observed change in behavioural variables due to

Table 4
Results for the diabetes-related biomarkers.

Health	Observed		Counterfactual 1		Counterfactual 2		Total effect		Non-behavioural effect		Behavioural effect	
	Estimates	SEs	Estimates	SEs	Estimates	SEs	Estimates	SEs	Estimates	SEs	Estimates	SEs
BMI												
Q05	18.32	0.04	18.36	0.16	18.46	0.18	−0.04	0.16	−0.15	0.17	0.10*	0.06
Q20	20.42	0.05	20.66	0.19	20.76	0.20	−0.24	0.19	−0.34*	0.20	0.10	0.07
Q35	21.76	0.05	22.14	0.22	22.31	0.23	−0.39*	0.22	−0.56***	0.23	0.17**	0.08
Q50	23.07	0.05	23.56	0.24	23.76	0.27	−0.49**	0.24	−0.69***	0.26	0.20**	0.09
Q65	24.44	0.04	24.97	0.27	25.21	0.30	−0.53*	0.27	−0.76***	0.29	0.23**	0.10
Q80	26.12	0.06	26.80	0.31	27.08	0.35	−0.67**	0.31	−0.96***	0.35	0.29**	0.12
Q95	29.37	0.10	30.26	0.41	30.69	0.44	−0.89**	0.39	−1.32***	0.43	0.42***	0.15
Mean	23.33	0.03	23.80	0.25	24.01	0.27	−0.47*	0.25	−0.68***	0.27	0.21***	0.09
Triglycerides												
Q05	46.94	0.52	44.80	1.71	45.73	1.78	2.14	1.66	1.22	1.72	0.92	0.59
Q20	69.09	0.64	64.58	2.84	66.24	3.05	4.51*	2.74	2.85	2.94	1.66*	0.97
Q35	89.46	0.82	81.97	3.90	84.29	4.23	7.49*	3.83	5.17	4.15	2.31*	1.34
Q50	110.72	0.94	102.15	5.21	105.69	5.65	8.57*	5.12	5.02	5.53	3.55**	1.74
Q65	141.72	1.52	128.57	7.23	133.23	7.89	13.15*	7.08	8.49	7.70	4.66*	2.54
Q80	194.86	2.47	173.83	10.86	180.86	12.10	21.04**	10.72	14.00	11.91	7.04*	3.91
Q95	348.98	5.89	306.90	22.76	323.33	26.40	42.08*	22.46	25.65	25.97	16.42*	9.78
Mean	143.32	1.21	129.43	7.72	134.48	8.69	13.89*	7.63	8.84	8.58	5.05*	2.83
HbA1c												
Q05	4.70	0.04	4.83	0.03	4.83	0.03	−0.13***	0.05	−0.13***	0.05	−0.00	0.01
Q20	5.10	0.04	5.27	0.04	5.27	0.04	−0.17***	0.05	−0.17***	0.05	0.01	0.02
Q35	5.30	0.00	5.48	0.04	5.48	0.04	−0.18***	0.04	−0.18***	0.04	0.01	0.02
Q50	5.50	0.00	5.66	0.05	5.67	0.05	−0.16***	0.05	−0.17***	0.05	0.00	0.02
Q65	5.70	0.00	5.86	0.05	5.87	0.05	−0.16***	0.05	−0.17***	0.05	0.02	0.02
Q80	5.90	0.01	6.12	0.06	6.13	0.06	−0.22***	0.06	−0.23***	0.06	0.01	0.02
Q95	6.70	0.06	6.91	0.11	6.94	0.11	−0.21**	0.10	−0.24**	0.11	0.02	0.04
Mean	5.60	0.01	5.77	0.05	5.78	0.06	−0.17***	0.05	−0.18***	0.05	0.01	0.02
Glucose												
Q05	75.42	0.23	72.44	0.59	72.36	0.60	2.98***	0.54	3.06***	0.56	−0.08	0.20
Q20	83.34	0.15	79.17	0.74	79.16	0.79	4.17***	0.72	4.18***	0.78	−0.01	0.26
Q35	87.97	0.17	83.00	0.86	83.08	0.94	4.97***	0.84	4.88***	0.92	0.08	0.31
Q50	91.98	0.21	86.65	1.03	86.66	1.09	5.33***	1.00	5.32***	1.08	0.01	0.34
Q65	97.11	0.25	90.62	1.21	90.56	1.28	6.49***	1.19	6.55***	1.26	−0.06	0.40
Q80	103.86	0.36	96.36	1.51	96.44	1.62	7.50***	1.47	7.42***	1.58	0.08	0.52
Q95	132.12	1.33	116.43	2.93	116.54	3.05	15.69***	2.70	15.58***	2.86	0.11	1.03
Mean	96.54	0.25	89.49	1.29	89.52	1.39	7.05***	1.26	7.02***	1.36	0.03	0.42

Note: Standard errors (SEs) are calculated by bootstrap with 500 repetitions.
The p-values come from testing the hypothesis that respective effects are zero.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

urban developments shifts the entire distribution of the HDL cholesterol towards the left, without changing the shape of the distribution.

For LDL cholesterol, we observe a significant positive total effect for all estimated percentiles ($p < 0.10$). The larger total effect is observed on the higher percentiles of the distribution, which implies that urbanicity has a stronger adverse effect among those people sitting at the higher percentiles of the distribution. We find that most parts of the total effect are attributable to the non-behavioural effect ($p < 0.10$) (Fig. 5), and the non-behavioural effect also shows a larger effect on the higher percentiles. The size of the behavioural effect is positive and constant across the distribution, but its large standard error makes the behavioural effect statistically insignificant.

Lastly, for total cholesterol, we find a significant positive effect ($p < 0.10$). We observe its larger effect on the higher percentiles (Fig. 5), but it also exhibits a larger standard error. While on the lower and middle percentiles of the distribution, the non-behavioural effect is larger than the behavioural effect, the latter becomes larger on the higher percentiles of the distribution. In fact, a greater proportion of the total effect on the higher percentiles is explained by the behavioural effect, which implies that a change in total cholesterol levels, due to the observed changes in behavioural variables, is more prominent among people with higher total cholesterol levels.

5. Discussion

China has experienced rapid urbanisation over the last few decades, but its implications for health are not straightforward. Despite a large number of studies on the subject, understanding of the implications for objective health outcomes is still limited. The use of information on various blood-based biomarkers, as well as anthropometric measurements and blood pressure, allowed us to explore the relationship between urban developments and various aspects of health.

This study quantifies the effect of urban developments on the distribution of objectively measured health. We quantify this effect by estimating a counterfactual health distribution in a situation in which everyone is exposed to the lowest level of urban development. We estimate the effect of urban developments on the entire health distribution beyond its mean, which helps elucidate the heterogeneous effect of urban developments across health distribution. Focusing on the entire distribution is important, because it is quite probable that factors influencing health at the bottom of distribution could be irrelevant at the top (Bitler, Gelbach, & Hoynes, 2006). Despite its importance, less research has been done so far, especially in developing countries (Aizawa, 2018b). Furthermore, we estimate the behavioural effect and the non-behavioural effect, the former of which is part of the total effect that can be attributable to the mechanism through which the urban environment affects observable health-related behaviours. The non-behavioural effect is the remaining part that cannot be explained by such mechanisms through observable behavioural change.

First, we find that urban developments affect considerably the distribution of observable behavioural variables, and a significant negative influence on physical activities is observed. A significant effect on dietary

Table 5
Results for the cardiovascular- and kidney-related biomarkers.

Health	Observed		Counterfactual 1		Counterfactual 2		Total effect		Non-behavioural effect		Behavioural effect	
	Estimates	SEs	Estimates	SEs	Estimates	SEs	Estimates	SEs	Estimates	SEs	Estimates	SEs
Systolic BP												
Q05	100.00	0.09	97.75	0.91	97.44	0.95	2.25***	0.91	2.56***	0.95	−0.32	0.34
Q20	110.00	0.00	107.30	1.09	107.19	1.17	2.70***	1.09	2.81***	1.17	−0.12	0.39
Q35	118.00	0.29	113.51	1.32	113.38	1.47	4.49***	1.31	4.62***	1.47	−0.13	0.47
Q50	120.67	0.08	117.71	1.29	117.57	1.36	2.95**	1.29	3.10**	1.36	−0.14	0.41
Q65	129.33	0.32	123.31	1.54	123.34	1.70	6.02***	1.53	5.99***	1.69	0.03	0.53
Q80	139.33	0.28	132.10	1.81	131.65	1.98	7.23***	1.80	7.68***	1.96	−0.45	0.57
Q95	160.00	0.17	150.47	2.31	150.49	2.44	9.53***	2.30	9.51***	2.44	0.02	0.73
Mean	124.68	0.20	120.03	1.42	119.90	1.54	4.65***	1.40	4.78***	1.52	−0.13	0.45
Diastolic BP												
Q05	63.33	0.44	61.83	0.79	61.58	0.84	1.50**	0.73	1.75**	0.78	−0.25	0.30
Q20	70.00	0.33	69.18	0.73	68.94	0.76	0.82	0.77	1.06	0.80	−0.25	0.25
Q35	76.67	0.32	75.09	0.89	74.81	1.01	1.58*	0.88	1.86*	1.00	−0.28	0.37
Q50	80.00	0.00	78.02	0.89	77.74	0.94	1.98**	0.89	2.26***	0.94	−0.29	0.31
Q65	82.00	0.20	80.37	0.99	80.44	1.05	1.63	1.00	1.56	1.05	0.07	0.39
Q80	89.33	0.14	86.73	1.06	86.35	1.15	2.61***	1.07	2.98***	1.15	−0.38	0.39
Q95	100.00	0.00	97.23	1.22	96.93	1.30	2.77**	1.22	3.07***	1.30	−0.30	0.44
Mean	80.33	0.12	78.42	0.92	78.24	0.99	1.91**	0.91	2.09**	0.98	−0.18	0.33
Uric acid												
Q05	2.99	0.02	2.96	0.07	3.06	0.07	0.03	0.07	−0.07	0.07	0.11***	0.03
Q20	3.80	0.02	3.77	0.09	3.88	0.09	0.03	0.08	−0.09	0.09	0.12***	0.03
Q35	4.40	0.02	4.35	0.10	4.49	0.10	0.06	0.10	−0.08	0.10	0.14***	0.03
Q50	4.97	0.02	4.92	0.11	5.07	0.12	0.06	0.11	−0.10	0.11	0.15***	0.04
Q65	5.58	0.02	5.49	0.13	5.69	0.13	0.09	0.13	−0.11	0.13	0.20***	0.05
Q80	6.39	0.03	6.28	0.15	6.50	0.16	0.11	0.15	−0.11	0.16	0.22***	0.06
Q95	8.03	0.05	7.92	0.19	8.17	0.21	0.11	0.19	−0.13	0.21	0.25***	0.08
Mean	5.15	0.02	5.09	0.11	5.26	0.12	0.06	0.11	−0.11	0.12	0.17***	0.04
Creatinine												
Q05	0.75	0.00	0.72	0.01	0.73	0.01	0.02***	0.01	0.02**	0.01	0.01***	0.00
Q20	0.84	0.00	0.80	0.01	0.81	0.01	0.04***	0.01	0.03***	0.01	0.01***	0.00
Q35	0.89	0.00	0.86	0.01	0.87	0.01	0.04***	0.01	0.03***	0.01	0.01***	0.00
Q50	0.96	0.00	0.91	0.01	0.92	0.01	0.05***	0.01	0.04***	0.01	0.01***	0.00
Q65	1.03	0.00	0.97	0.01	0.98	0.01	0.06***	0.01	0.05***	0.01	0.02***	0.00
Q80	1.12	0.00	1.04	0.01	1.06	0.01	0.08***	0.01	0.06***	0.01	0.02***	0.00
Q95	1.29	0.01	1.18	0.02	1.21	0.02	0.10***	0.02	0.08***	0.02	0.02***	0.01
Mean	0.98	0.00	0.93	0.01	0.94	0.01	0.06***	0.01	0.04***	0.01	0.01***	0.00

Note: Standard errors (SEs) are calculated by bootstrap with 500 repetitions.
The p-values come from testing the hypothesis that respective effects are zero.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

patterns is also observed, in that a dietary change from traditional to Western foods is a by-product of economic development and urbanisation (Popkin, 2001; Popkin & Du, 2003; Popkin et al., 2006; Popkin & Gordon-Larsen, 2004), and it may have been accelerated by a change in food supply via urbanisation; for example, the trend for fresh markets as major sources of food in developing countries is being replaced by multinational, large and regional supermarkets providing processed higher-fat, added-sugar and salt-laden foods (Hu, Reardon, Rozelle, Timmer, & Wang, 2004; Popkin et al., 2006; Reardon & Berdegué, 2002), which may have contributed to the shift in dietary patterns. As well as the changes in physical activities and dietary patterns, we also find that urban developments are related closely to higher incomes and greater educational achievement, as well as to occupational change from farming-based industry to the service industry.

The main results of this study reveal the significant total positive effects of urbanicity on the means of triglycerides, glucose, systolic blood pressure, diastolic blood pressure, creatinine, HDL cholesterol, LDL cholesterol and total cholesterol. The main results also provide strong evidence of the existence of heterogeneity in the effects' sizes across the distribution, which in turn means that the effect of urban developments is not constant across the distribution, and heterogeneous effects change the shape of the distribution of health outcomes as well as shifting its location. The larger total effects found at the higher percentiles of triglycerides, glucose, systolic blood pressure, diastolic blood pressure, creatinine, HDL cholesterol, LDL cholesterol and total cholesterol imply that people sitting at the higher percentiles of the respective distributions are more responsive to urban environments than those who are sitting at the lower and middle points. These heterogeneities across the distribution are suggestive of changes in the public health risk structure through urban development. Overall, the results indicate that urban developments have adverse effects especially on biomarkers related to body lipids, such as triglycerides and cholesterols, and biomarkers related to cardiovascular and kidney-related diseases.

Decomposing the total effects into behavioural and non-behavioural effects, we find that the observed behavioural changes as a result of urbanicity contribute significantly to the change in the biomarkers' distributions. Especially, we observe that the behavioural effect has a larger effect on the higher percentiles of the distributions of BMI, triglycerides, uric acid, creatinine and total cholesterol. These heterogeneous effects across the distribution tell us that people sitting at the higher percentiles thereof are more subject to the behavioural changes than those who are sitting at the lower and middle percentiles. On the other hand, the behavioural effect is relatively stable or small across the distributions of HbA1c, glucose, blood pressure, HDL cholesterol and LDL cholesterol, which suggests that, for the case of diabetes-related biomarkers, cardiovascular-related biomarkers and cholesterols, observed behavioural changes through urban development shift the location of their distributions, without changing their shapes.

With regard to the health outcomes for which behavioural effects show large contributions to their corresponding total effects, policymakers in China need to think about policies, in order to mitigate partially the adverse effects of urbanicity by changing health-related behaviours among city dwellers and by promoting healthy lifestyles. For example, in terms of food consumption, enhancing nutritional

Table 6
Results for the cholesterols.

Health	Observed		Counterfactual 1		Counterfactual 2		Total effect		Non-behavioural effect		Behavioural effect	
	Estimates	SEs	Estimates	SEs	Estimates	SEs	Estimates	SEs	Estimates	SEs	Estimates	SEs
HDL cholesterol												
Q05	35.19	0.27	33.23	0.78	32.65	0.86	1.96***	0.77	2.54***	0.85	−0.58*	0.33
Q20	42.92	0.20	40.68	0.94	40.03	0.97	2.24***	0.94	2.89***	0.97	−0.65*	0.36
Q35	48.34	0.23	45.49	0.98	44.83	1.01	2.84***	0.98	3.51***	1.02	−0.67*	0.36
Q50	53.36	0.21	50.25	1.08	49.56	1.13	3.12***	1.08	3.81***	1.13	−0.69*	0.40
Q65	58.78	0.24	55.13	1.14	54.33	1.19	3.65***	1.14	4.45***	1.19	−0.79**	0.40
Q80	66.13	0.29	62.05	1.31	61.24	1.35	4.07***	1.29	4.89***	1.34	−0.82*	0.47
Q95	80.82	0.49	76.61	1.60	75.90	1.63	4.21***	1.64	4.92***	1.68	−0.71	0.58
Mean	55.01	0.16	51.90	1.11	51.17	1.15	3.11***	1.10	3.84***	1.15	−0.73*	0.39
LDL cholesterol												
Q05	63.03	0.75	59.01	2.04	59.62	2.14	4.02**	1.90	3.41*	2.00	0.61	0.76
Q20	85.85	0.50	80.49	2.26	81.15	2.36	5.35***	2.22	4.70**	2.32	0.65	0.78
Q35	100.15	0.53	94.67	2.49	95.02	2.61	5.49**	2.46	5.14**	2.58	0.35	0.86
Q50	112.92	0.51	106.69	2.65	107.37	2.78	6.23***	2.65	5.54**	2.79	0.68	0.96
Q65	126.06	0.53	119.56	2.83	120.19	2.92	6.50**	2.84	5.88**	2.92	0.62	0.96
Q80	142.69	0.58	136.25	3.11	136.93	3.28	6.44**	3.13	5.76*	3.29	0.68	1.06
Q95	176.89	1.03	169.91	3.60	170.38	3.72	6.99*	3.63	6.52*	3.74	0.47	1.30
Mean	115.17	0.39	109.26	2.66	109.88	2.79	5.91**	2.66	5.29*	2.79	0.61	0.91
Total cholesterol												
Q05	133.02	0.74	129.29	1.97	131.00	2.11	3.74*	1.94	2.02	2.04	1.72**	0.83
Q20	156.23	0.54	151.84	2.33	153.70	2.47	4.38*	2.24	2.53	2.37	1.86**	0.87
Q35	171.31	0.59	166.09	2.56	168.27	2.72	5.22**	2.48	3.04	2.64	2.17**	0.96
Q50	184.84	0.53	179.58	2.76	181.84	3.02	5.26*	2.73	3.00	2.98	2.26**	1.02
Q65	199.54	0.62	193.73	3.06	196.40	3.28	5.80*	3.08	3.13	3.27	2.67***	1.13
Q80	218.48	0.66	212.98	3.38	215.56	3.70	5.50	3.39	2.92	3.68	2.58**	1.28
Q95	254.83	1.18	248.57	4.13	252.39	4.55	6.27	4.18	2.44	4.58	3.82***	1.59
Mean	188.07	0.42	183.03	2.81	185.32	3.02	5.04*	2.79	2.76	3.00	2.29**	1.02

Note: Standard errors (SEs) are calculated by bootstrap with 500 repetitions.
The p-values come from testing the hypothesis that respective effects are zero.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

knowledge and further food regulations, such as promoting nutrition labels for processed foods and restaurant menus, seems important.[20] Economic incentives such as the taxation of energy-dense foods, sugar-added beverages and processed foods high in trans fats could be another option. These policies are becoming increasingly important as dietary patterns shift rapidly towards Westernised diets, hand in hand with the rise in household income and the choice to eat out becoming popular among city dwellers (Popkin & Du, 2003; Guo et al., 2000; Drewnowski, 2000; Du et al., 2002, 2004). Enhancing "health-related literacy" through education would become more important and promising given the finding that urban developments offer better educational opportunities for a wide swathe of the population.

Of course, diet is a major consensus-based contributor to the risk of NCDs, but multidimensional health policies in collaboration with multi-sectoral industries will also be necessary to address effectively the rapid increase in NCD risk factors. The adverse effect of sedentary urban lifestyles on health should also be re-acknowledged. Creating healthy urban environments that allow us to have healthy lifestyles is therefore of paramount importance. For example, urban designs promoting more physical exercise – required to tackle greater inactivity – would be beneficial. Furthermore, encouraging the choice of active modes of transport, such as walking, cycling and public transport, and developing transport infrastructure are promising actions in this regard (Cervero & Gorham, 1995). This perspective is becoming more and more important in light of the shift of occupational choice from the farming industry to less labour intensive service industry, in which case policymakers in China may therefore wish to consider re-acknowledging the importance of incorporating urban health perspectives in promoting urban development.

[20] For South Korea, for instance, the promotion of a traditional diet which is low in fat and rich in vegetables has been successful in the fight against the increasing risks of NCDs (Kim, Moon, & Popkin, 2000; Lee, Popkin, & Kim, 2002).

6. Conclusion

By way of conclusion, it is worth touching upon a few limitations and potential extensions of this study. First, thanks to the comprehensive information provided by the CHNS, this study includes a wide range of health-related behavioural variables, including nutrition intake and physical activities, but health-related behaviours that could be influenced by urban developments are without doubt far more complicated. For instance, urbanisation can change the characteristics of the community; urban sprawl, social isolation and the development of slums can considerably alter the human relationships between dwellers (Wang, Wang, & Wu, 2009). Although the community effect and the peer effect could be considered potentially important health-related behavioural variables of health (Kawachi, Kennedy, Lochner, & Prothrow-Stith, 1997; Villalonga-Olives & Kawachi, 2017), they are usually unobservable and hardly quantifiable. If we were able to measure these factors, however, it could help us grasp better the underlying mechanism behind how urban developments influence health. Exploring the behavioural changes caused by these psychological factors would complement the findings of this study.

Second, in the decomposition analysis, we could not explore what behavioural variable(s) are associated strongly with the estimated behavioural effect. As we decomposed the observed total effects at various quantile points of the distribution beyond the mean, decomposing further the overall behavioural effect into the parts attributable to individual behavioural variables is not technically straightforward, because, in contrast to the case of the mean, the law of iterated expectation cannot be applied for quantiles (Fortin, Lemieux, & Firpo, 2011). Exploring the contributions made by each individual behavioural variable with an improved robust statistical method would be our future challenge.

Third, this study is a cross-sectional analysis, and hence the effects measured herein are a mixture of temporal and cumulative effects. Disentangling them, by exploiting longitudinal data and analysing the

lifetime effect on health of exposure to the urban environment, would be a promising research topic. Lastly, another extension would be to see whether similar results can be found in other countries in Asia and possibly beyond. Many developing countries are currently urbanising while at the same time witnessing the rapid spread of NCDs. New empirical research in other countries on this topic should therefore be of paramount interest to public health specialists, urban planners and policymakers around the world.

Ethics approval

This research does not use data that require ethics approval. The research did not implement any experiments to humans and did not collect data by myself. This research uses public data without identifiable private information such as names or addresses of domicile of respondents.

Conflicts of interest

There are no known conflicts of interest associated with this publication. There is no financial support for this study.

A. Appendix

Table A.1
Cut-off values.

	Lower cutoff	Higher cutoff	Number of dropped observations
BMI (kg/m^2)	16.02	33.73	92
Triglyceride (mg/dL)	31.89	919.40	81
HbA1c ($mmol/L$)	4.00	10.90	74
Glucose (mg/dL)	62.69	265.84	84
Systolic blood pressure ($mmHg$)	89.33	189.33	74
Diastolic blood pressure ($mmHg$)	56.67	120.00	54
Uric acid (mg/dL)	2.17	11.60	84
Creatinine (mg/dL)	0.64	1.81	72
HDL Cholesterol (mg/dL)	25.52	116.40	80
LDL Cholesterol (mg/dL)	27.84	233.95	84
Total cholesterol (mg/dL)	107.50	317.09	82

Note: Cut-off points are defined by the top 0.1% and bottom 0.1% of each health outcome.

Table A.2
Mean comparison between the excluded and included samples

	Excluded observations		Included observations		Difference		
	N	Mean	N	Mean	Estimate	Standard error	p value
BMI (kg/m^2)	582	23.35	8625	23.33	0.02	0.14	0.89
Triglyceride (mg/dL)	539	155.4	7856	143.32	12.08	4.82	0.01
HbA1c ($mmol/L$)	536	5.65	7821	5.6	0.04	0.03	0.18
Glucose (mg/dL)	538	97.9	7854	96.54	1.36	0.98	0.16
Systolic blood pressure ($mmHg$)	532	125.16	7631	124.68	0.48	0.81	0.55
Diastolic blood pressure ($mmHg$)	531	80.55	7651	80.33	0.22	0.48	0.65
Uric acid (mg/dL)	541	5.2	7849	5.15	0.04	0.07	0.52
Creatinine (mg/dL)	542	0.98	7860	0.98	0.00	0.01	0.93
HDL Cholesterol (mg/dL)	537	54.43	7858	55.01	−0.58	0.63	0.35
LDL Cholesterol (mg/dL)	537	114.13	7852	115.17	−1.04	1.54	0.50
Total cholesterol (mg/dL)	539	187.96	7855	188.07	−0.12	1.64	0.94

Table A.3
Descriptive statistics stratified by urban/rural

	Urban					Rural				
	count	mean	sd	min	max	count	mean	sd	min	max
BMI	5737	23.27	3.30	16.02	33.73	2888	23.46	3.35	16.03	33.72
Triglyceride (mg/dL)	5251	140.62	103.94	31.89	919.40	2605	148.74	114.42	31.89	913.20
HbA1c (mmol/L)	5230	5.59	0.73	4.00	10.90	2591	5.64	0.78	4.00	10.90
Glucose (mg/dL)	5249	96.27	21.21	62.69	258.12	2605	97.07	22.90	64.08	265.84
Systolic BP (mmHg)	5141	124.65	18.17	89.33	189.33	2490	124.75	17.73	89.33	188.00
Diastolic BP (mmHg)	5158	80.31	11.02	56.67	120.00	2493	80.36	10.25	56.67	120.00
Uric Acid (mg/dL)	5253	5.09	1.52	2.17	11.51	2596	5.28	1.63	2.17	11.60
Creatinine (mg/dL)	5249	0.98	0.17	0.64	1.81	2611	1.00	0.18	0.64	1.78
HDL cholesterol (mg/dL)	5254	55.50	14.07	25.52	116.40	2604	54.02	13.84	25.52	114.85
LDL cholesterol (mg/dL)	5241	114.56	34.23	27.84	233.95	2611	116.39	34.98	28.23	233.95
Total cholesterol (mg/dL)	5243	187.64	37.21	107.50	317.09	2612	188.93	35.94	107.89	316.71

(*continued on next page*)

Table A.3 (*continued*)

	Urban					Rural				
	count	mean	sd	min	max	count	mean	sd	min	max
Urbanicity index	6049	61.22	17.49	30.42	103.07	3156	80.16	16.79	37.80	106.46
Age 30-39	6049	0.16	0.37	0.00	1.00	3156	0.14	0.35	0.00	1.00
Age 40-49	6049	0.24	0.43	0.00	1.00	3156	0.21	0.41	0.00	1.00
Age 50-59	6049	0.23	0.42	0.00	1.00	3156	0.25	0.43	0.00	1.00
Age over 60	6049	0.28	0.45	0.00	1.00	3156	0.30	0.46	0.00	1.00
Male	6049	0.48	0.50	0.00	1.00	3156	0.47	0.50	0.00	1.00
Urban registration	6049	0.27	0.44	0.00	1.00	3156	0.70	0.46	0.00	1.00
Married	6049	0.85	0.36	0.00	1.00	3156	0.82	0.39	0.00	1.00
Family size	6049	3.14	1.56	1.00	10.00	3156	3.10	1.30	1.00	8.00
Professional worker	6049	0.06	0.23	0.00	1.00	3156	0.10	0.31	0.00	1.00
Farmer	6049	0.37	0.48	0.00	1.00	3156	0.07	0.26	0.00	1.00
Self-employed/business owner	6049	0.46	0.50	0.00	1.00	3156	0.13	0.34	0.00	1.00
Permanent worker	6049	0.10	0.30	0.00	1.00	3156	0.17	0.37	0.00	1.00
Contractor	6049	0.04	0.18	0.00	1.00	3156	0.08	0.28	0.00	1.00
Temporary worker	6049	0.05	0.22	0.00	1.00	3156	0.06	0.24	0.00	1.00
Family size	6049	3.14	1.56	1.00	10.00	3156	3.10	1.30	1.00	8.00
Education years	6049	6.63	4.50	0.00	16.00	3156	8.56	4.85	0.00	16.00
Household income (divided by 1,000)	6049	8.75	8.59	0.00	77.92	3156	13.19	10.71	0.00	71.37
Energy intake (kcal)	6049	2160.43	619.13	587.55	4176.46	3156	1999.12	563.04	639.90	4188.97
Fat intake (g)	6049	71.54	33.45	8.24	209.66	3156	76.92	34.40	8.23	206.95
Protein intake (g)	6049	64.63	21.40	16.91	143.36	3156	66.57	21.78	18.19	142.62
Carbohydrate intake (g)	6049	309.40	97.65	76.00	615.03	3156	255.44	87.38	79.22	613.84
METs	6049	128.14	125.31	0.00	558.51	3156	101.15	103.64	0.00	546.25

Age, sex, registarion, marital status, and occupation types are binary variables.

Table A.4
Regression results for behavioural variables

Box-Cox model	Household income	Kcal	Protein	Carbohydrate	Fat	METs	Education	Family size
Coefficients								
Urban index	0.099***	0.89***	0.034	0.22***	0.16***	0.26**	−0.07	−0.067***
Age 30-39	−0.091	4.3***	0.19	0.95*	1.1**	0.42	−1.3***	−0.22***
Age 40-49	0.38*	5.4***	0.33**	1.4***	0.91*	1.3	−1.4***	−0.69***
Age 50-59	−0.16	4.2***	0.1	0.69	1.2***	1.5*	−3.5***	−0.59***
Age over 60	−0.88***	0.1	−0.38***	0.1	0.14	−2.2***	−4.8***	−0.6***
Male	0.075	6.7***	0.53***	1.6***	0.8***	1.1***	1.9***	0.0063
Urban hukou	−0.2	−0.89	0.27*	−1.4***	0.73	−6.4***	1.8***	0.36***
Uindex squared	−0.0015***	−0.013***	−0.00041	−0.0039***	−0.0016	−0.0034**	0.001	0.00095***
Uindex cubed	0.0000074***	0.00006***	0.000002	0.000018***	0.0000055	0.000012	−0.0000032	−0.0000042***
Age 30–39:Uindex	0.00006	−0.051**	−0.0023	−0.0095	−0.015**	0.0058	0.0071**	0.0029***
Age 40–49:Uindex	−0.0033	−0.058**	−0.0048**	−0.017***	−0.0089	−0.015	0.0064*	0.0048***
Age 50–59:Uindex	0.004	−0.046*	−0.002	−0.0079	−0.014**	−0.037***	0.018***	0.0029***
Age over 60:Uindex	0.012***	−0.041*	−0.00087	−0.013**	−0.0068	−0.038***	0.014***	0.0028***
Male:Uindex	0.00012	−0.0055	−0.000054	−0.00099	0.00086	−0.014***	−0.012***	0.0003
Urban hukou:Uindex	0.013***	0.021	−0.0026	0.018***	−0.0043	0.078***	−0.0082*	−0.0066***
Box-cox parameter								
Theta	0.27***	0.48***	0.29***	0.4***	0.44***	0.31***	0.64***	0.034**

Logit model	Married	Professional	Farmer	Self-employed	Permanent	Contractor	Temporary
Coefficients							
Urban index	0.032	0.03	0.096	0.14***	0.14***	−0.024	0.11**
Age 30-39	1.1***	0.12	0.37	0.94***	−0.58**	−0.74**	−0.094
Age 40-49	1.4***	−0.13	0.35	1.2***	−1.1***	−0.67**	−0.15
Age 50-59	0.63***	0.15	0.0071	1.3***	−0.73***	−0.65**	−0.58*
Age over 60	−0.35*	0.061	−1***	0.61***	−0.92**	−1.3***	−0.75**
Male	−0.17	0.43**	0.27	0.24**	0.32*	0.53***	0.91***
Urban hukou	0.5*	1.2***	−2.5***	−2.9***	1.2***	0.79**	0.27
Uindex squared	−0.00077	−0.0004	−0.0017	−0.0021***	−0.002***	0.00067	−0.0011
Uindex cubed	0.0000042	0.0000024	0.0000047	0.0000086***	0.0000095***	−0.0000039	0.0000039
Age 30–39:Uindex	0.0056	−0.0015	0.0032	−0.0062*	0.0088**	0.0064	0.001
Age 40–49:Uindex	0.0052	0.0022	0.0068	−0.01***	0.016***	0.0029	0.0016
Age 50–59:Uindex	0.013***	−0.0038	0.014*	−0.014***	0.005	−0.0022	0.0033
Age over 60:Uindex	0.016***	−0.016***	0.026***	−0.012***	−0.0094*	−0.003	−0.00053
Male:Uindex	0.0048***	−0.00079	−0.0035	−0.000031	0.0019	−0.0052**	−0.0091***
Urban hukou:Uindex	−0.0082**	−0.0056	0.015**	0.029***	−0.0062	−0.0056	−0.0058

Standard errors (SEs) are calculated by bootstrap with 500 repetitions (not shown here).
The P-values come from testing the hypothesis that respective coefficients and a transformation parameter are zero.
*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Table A.5
Regression results for health outcomes

	BMI	Triglycerides	HbA1c	Glucose	Systolic BP	Diastolic BP	Uric acid	Creatinine	HDL cholesterol	LDL cholesterol	Total cholesterol
Coefficients											
Urban index	−0.00038	0.00078	−0.00014	0.0000017***	0.000099	0.0023	−0.017**	−0.00057	0.018***	−0.1	−0.01
Age 30-39	0.049***	−0.018	0.00026	0.0000053*	0.00013	0.012	0.0051	0.033	0.011	1	0.09*
Age 40-49	0.045***	−0.0041	0.0016**	0.000008***	0.0006	0.027**	−0.11**	0.0079	0.085*	2.1***	0.15***
Age 50-59	0.033*	0.032	0.002***	0.000012***	0.00099**	0.022**	−0.019	−0.0039	0.08	2.8***	0.24***
Age over 60	−0.014	−0.019	0.0018***	0.00001***	0.0013***	0.026**	−0.028	0.044*	0.19***	3***	0.26***
Male	−0.039***	−0.046***	0.00032	0.00000041	−0.000097	0.0067	0.27***	0.16***	0.031	−0.5	−0.061**
Urban hukou	0.085***	0.035	0.0004	0.0000035	−0.00036	−0.0021	0.021	0.011	−0.07	0.46	0.053
Household income	−0.000023	−0.000022	0.0000017	−0.000000021***	−0.0000016*	−0.000041*	0.00021*	0.000025	−0.00016	0.0021	0.00021*
Kcal	0.00000017	0.00000013	−0.0000000075	0.000000000091**	0.000000007	0.00000019*	−0.00000089	−0.00000016	0.00000049	−0.000013*	−0.0000012**
Protein	−0.016*	−0.023	−0.00077*	−0.0000015	−0.00009	−0.0022	−0.06**	−0.0011	0.023	0.3	0.01
Carbohydrate	0.026**	0.012	0.00093*	0.0000015	−0.0004	−0.0074	0.01	0.021	−0.011	0.091	−0.022
Fat	0.026	0.031	0.0013	−0.0000042	0.00079	0.034***	0.11	−0.028	−0.085	0.74	0.0039
METs	−0.017	−0.0093	0.001	−0.0000073*	0.00021	0.0029	−0.12*	−0.086***	0.059	−1.1	−0.099*
Education	0.017	0.0004	−0.0009	0.00000094	−0.0002	0.00084	0.077	0.086***	−0.037	1.7**	0.12**
Family size	−0.013	0.051	−0.00047	0.0000013	0.00034	0.01	−0.013	0.066**	−0.093	0.21	0.056
Married	−0.023	0.0084	−0.0007	0.0000014	−0.0002	−0.0067	0.094	0.11***	0.033	0.7	0.046
Professional	−0.039**	0.0061	−0.0012	−0.0000038	−0.000021	−0.0045	0.038	0.054*	−0.064	−0.09	−0.046
Farmer	0.0014	0.0021	−0.0003***	−0.0000011*	−0.000074	−0.00017	0.0048	0.000098	0.0049	−0.057	0.0034
Self-employed	−0.0012	−0.007*	−0.000088	0.000001*	0.00012*	0.0019	0.00034	−0.0046	0.026***	0.022	0.0089
Permanent	0.0083	0.022**	0.0012***	0.00000056	−0.00021	−0.0044	0.025	0.0039	−0.03	−0.12	−0.0049
Contractor	−0.00097	0.012	0.000016	−0.0000027**	−0.00027**	−0.0043	−0.016	0.0063	−0.057***	−0.15	−0.03**
Temporary	0.0023	0.0083	−0.000078	−0.0000022***	−0.00014	−0.0035	0.014	0.013**	−0.043***	−0.011	−0.0092
Uindex squared	−0.00054	0.000057	0.0001***	0.00000015	−0.000022	−0.00015	−0.0032	−0.0026**	0.002	−0.027	−0.00034
Uindex cubed	0.0018	−0.0021	−0.00011	0.00000013	0.000072*	0.00032	−0.0057	−0.0016	0.0053	0.12*	0.0054
Age 30–39:Uindex	−0.00022	0.00074*	0.0000041	−0.0000000094	0.0000047	−0.000031	−0.00024	−0.00016	−0.00024	−0.0019	−0.00012
Age 40–49:Uindex	0.00016	0.0011***	−0.0000028	0.0000000049	0.0000084	−0.000036	0.0014*	0.00032	−0.0014**	−0.0031	−0.000052
Age 50–59:Uindex	0.00029	0.00066	0.0000024	−0.000000033	0.000011*	0.00013	0.0006	0.00062*	−0.001	−0.003	−0.00049
Age over 60:Uindex	0.00077***	0.0011***	0.000017	0.000000013	0.000017***	0.000048	0.0012	0.00055	−0.0023***	−0.0016	−0.00073
Male:Uindex	0.00055***	0.001***	−0.0000028	0.000000017	0.0000074**	0.000079	0.00043	0.00055***	−0.0016***	0.0015	0.00059
Urban hukou:Uindex	−0.00097***	−0.00049	0.0000047	−0.000000036	0.0000036	0.000048	−0.00025	−0.00029	0.00095	−0.0049	−0.00065
Income:Uindex	0.00016	0.00019	0.0000082	0.000000005	−0.00000047	0.000024	0.00073*	0.00013	−0.00038	−0.0048	−0.00028
Kcal:Uindex	−0.00021	−0.00026	−0.0000056	−0.0000000092	0.0000015	0.000091	−0.0006	−0.00053*	0.0001	−0.0039	0.000047
Protein:Uindex	−0.00037	−0.00013	−0.000018	0.000000068	−0.0000099	−0.0004***	−0.00063	0.00054	0.001	−0.01	0.000071
Carbo:Uindex	0.000021	−0.000053	−0.000023	0.00000012*	−0.0000041	−0.000083	0.0016	0.0012***	−0.00033	0.011	0.0012
Fat:Uindex	−0.00017	−0.000096	0.0000058	−0.000000041	0.00000064	−0.000032	−0.0012	−0.001***	0.00046	−0.019*	−0.0014*
METs:Uindex	0.00014	−0.00082*	0.0000068	−0.000000026	−0.0000057	−0.00016	−0.0006	−0.00091**	0.0012	−0.0029	−0.00067
Education:Uindex	0.0002	−0.0004	0.000008	−0.000000032	0.0000023	0.000074	−0.0013	−0.0013***	−0.0003	−0.013	−0.00088
Family size:Uindex	0.00034	−0.00044	0.0000094	0.000000012	−0.0000025	0.000044	−0.0011	−0.00078*	0.0012	−0.0023	0.00023
Married:Uindex	0.00002	−0.000013	0.0000054***	0.000000017**	0.0000014	0.0000046	0.000019	0.000028	−0.000078	0.00041	−0.000076
Professional:Uindex	0.000015	0.00011*	0.0000012	−0.0000000013	−0.0000011	−0.000016	0.000074	0.000033	−0.00019*	−0.0006	−0.000063
Farmer:Uindex	−0.000039	−0.00029*	−0.000012***	−0.0000000046	0.0000027	0.000072	−0.0002	−0.000062	0.00024	0.0034	0.00018
Self-employed:Uindex	0.000022	−0.0002	−0.0000012	0.000000031*	0.0000027	0.000038	0.0000057	−0.000023	0.00038*	0.0022	0.00021
Permanent:Uindex	−0.000033	−0.00015	0.00000035	0.000000028**	0.0000011	0.000023	−0.00031*	−0.000083	0.00028	0.00055	0.0000023
Contractor:Uindex	0.000011	−0.000011	−0.0000015***	−0.0000000021	0.00000033	0.0000039	0.000052	0.000014	−0.000017	0.00044	0.0000031
Temporary:Uindex	−0.000038	0.000021	0.0000011	−0.0000000044	−0.0000016**	−0.000012	0.000032	0.0000028	−0.00011	−0.00082	−0.000051
Box-cox parameter											
Theta	−0.15***	−0.28***	−1.9***	−2***	−0.87***	−0.26***	0.062**	−0.74***	0.0044	0.52***	0.045

Coefficients are estimated by the Box-Cox model.
Standard errors (SEs) are calculated by bootstrap with 500 repetitions (not shown here).
The P-values come from testing the hypothesis that respective coefficients and a transformation parameter are zero.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Tudor-Locke, C., et al. (2011). 2011 compendium of physical activities: A second update of codes and met values. *Medicine & Science in Sports & Exercise, 43*(8), 1575–1581.

Ainsworth, B., Haskell, W., Leon, A., Jacobs, D., Montoye, H., Sallis, J., et al. (1993). Compendium of physical activities: Classification of energy costs of human physical activities. *Medicine & Science in Sports & Exercise, 25*(1), 71–80.

Aizawa, T. (2018a). Inequality in the treatment of diabetes and hypertension across residency status in China. *Ethnicity and Health*. https://doi.org/10.1080/13557858.2018.1539220.

Aizawa, T. (2018b). Regional disparity in the bmi distribution of Indonesian people: New evidence "beyond the mean". *Bulletin of Indonesian Economic Studies, 54*(1), 85–112.

Aizawa, T. (2019). Transition of the bmi distribution in India: Evidence from a distributional decomposition analysis. *Journal of Bioeconomics, 21*(1), 3–36.

Bell, A., Ge, K., & Popkin, B. (2001). Weight gain and its predictors in Chinese adults. *International Journal of Obesity, 25*, 1079–1086.

Bell, A. C., Ge, K., & Popkin, B. M. (2002). The road to obesity or the path to prevention: Motorized transportation and obesity in China. *Obesity Research, 10*(4), 277–283.

Benzeval, M., Kumari, M., & Jones, A. M. (2016). How do biomarkers and genetics contribute to understanding society? *Health Economics, 25*(10), 1219–1222.

Bitler, M. P., Gelbach, J. B., & Hoynes, H. W. (2006). What mean impacts miss: Distributional effects of welfare reform experiments. *The American Economic Review, 96*(4), 988–1012.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological), 26*(2), 211–252.

Carrieri, V., & Jones, A. M. (2017). The income–health relationship 'beyond the mean': New evidence from biomarkers. *Health Economics, 26*(7), 937–956.

Cervero, R., & Gorham, R. (1995). Commuting in transit versus automobile neighborhoods. *Journal of the American Planning Association, 61*(2), 210–225.

Chan, K. W., & Zhang, L. (1999). The hukou system and rural-urban migration in China: Processes and changes. *The China Quarterly, 160*, 815–855.

Chen, J. (1994). Dietary transition in China and its health consequences. *Asia Pacific Journal of Clinical Nutrition, 3*(3), 111–114.

Chen, J. (2007). Rapid urbanization in China: A real challenge to soil protection and food security. *Catena, 69*(1), 1–15.

Chen, H., Liu, Y., Zhu, Z., & Li, Z. (2017). Does where you live matter to your health? Investigating factors that influence the self-rated health of urban and rural Chinese residents: Evidence drawn from Chinese general social survey data. *Health and Quality of Life Outcomes, 15*(1), 78.

Cyril, S., Oldroyd, J. C., & Renzaho, A. (2013). Urbanisation, urbanicity, and health: A systematic review of the reliability and validity of urbanicity scales. *BMC Public Health, 13*(1), 513.

Dahly, D. L., & Adair, L. S. (2007). Quantifying the urban environment: A scale measure of urbanicity outperforms the urban–rural dichotomy. *Social Science & Medicine, 64*(7), 1407–1419.

Dans, A., Ng, N., Varghese, C., Tai, E. S., Firestone, R., & Bonita, R. (2011). The rise of chronic non-communicable diseases in southeast asia: Time for action. *The Lancet, 377*(9766), 680–689.

Deaton, A. (2013). *The great escape: Health, wealth, and the origins of inequality*. Princeton, NJ: Princeton University Press.

Drewnowski, A. (2000). Nutrition transition and global dietary trends. *Nutrition, 16*(7–8), 486–487.

Du, S., Lu, B., Zhai, F., & Popkin, B. M. (2002). A new stage of the nutrition transition in China. *Public Health Nutrition, 5*(1a), 169–174.

Du, S., Mroz, T. A., Zhai, F., & Popkin, B. M. (2004). Rapid income growth adversely affects diet quality in China? Particularly for the poor!. *Social Science & Medicine, 59*(7), 1505–1515.

Dye, C. (2008). Health and urban living. *Science, 319*(5864), 766–769.

Fang, H., Chen, J., & Rizzo, J. A. (2009). Explaining urban-rural health disparities in China. *Medical Care, 47*(12), 1209–1216.

Fortin, N. M., Lemieux, T., & Firpo, S. (2011). Decomposing methods in economics. In O. Ashenfelter, & D. Card (Eds.). *Handbook of labor economics, Volume Vol. 4a* (pp. 1–102). North Holland: Elsevier.

French, S. A., Story, M., & Jeffery, R. W. (2001). Environmental influences on eating and physical activity. *Annual Review of Public Health, 22*(1), 309–335.

Galea, S., & Vlahov, D. (2005). Urban health: Evidence, challenges, and directions. *Annual Review of Public Health, 26*(1), 341–365.

Gong, P., Liang, S., Carlton, E. J., Jiang, Q., Wu, J., Wang, L., et al. (2012). Urbanisation and health in China. *The Lancet, 379*(9818), 843–852.

Gorber, S. C., Tremblay, M., Moher, D., & Gorber, B. (2007). A comparison of direct vs. self-report measures for assessing height, weight and body mass index: A systematic review. *Obesity Reviews, 8*(4), 307–326.

Guo, X., Mroz, T. A., Popkin, B. M., & Zhai, F. (2000). Structural change in the impact of income on food consumption in China, 1989–1993. *Economic Development and Cultural Change, 48*(4), 737–760.

Hu, D., Reardon, T., Rozelle, S., Timmer, P., & Wang, H. (2004). The emergence of supermarkets with Chinese characteristics: Challenges and opportunities for China's agricultural development. *Development Policy Review, 22*(5), 557–586.

Jones-Smith, J. C., & Popkin, B. M. (2010). Understanding community context and adult health changes in China: Development of an urbanicity scale. *Social Science & Medicine, 71*(8), 1436–1446.

Kawachi, I., Kennedy, B. P., Lochner, K., & Prothrow-Stith, D. (1997). Social capital, income inequality, and mortality. *American Journal of Public Health, 87*(9), 1491–1498 PMID: 9314802.

Kim, Y. J. (2016). The long-run effect of education on obesity in the us. *Economics and Human Biology, 21*, 100–109.

Kim, S., Moon, S., & Popkin, B. M. (2000). The nutrition transition in South Korea. *American Journal of Clinical Nutrition, 71*(1), 44–53.

Konteh, F. H. (2009). Urban sanitation and health in the developing world: Reminiscing the nineteenth century industrial nations. *Health & Place, 15*(1), 69–78.

Lee, M. J., Popkin, B. M., & Kim, S. (2002). The unique aspects of the nutrition transition in South Korea: The retention of healthful elements in their traditional diet. *Public Health Nutrition, 5*(1a), 197–203.

Leon, D. A. (2008). Cities, urbanization and health. *International Journal of Epidemiology, 37*(1), 4–8.

Li, X., Song, J., Lin, T., Dixon, J., Zhang, G., & Ye, H. (2016). Urbanization and health in China, thinking at the national, local and individual levels. *Environmental Health, 15*(32), 113–123.

Liu, Z. (2005). Institution and inequality: The hukou system in China. *Journal of Comparative Economics, 33*(1), 133–157.

Liu, G. G., Wu, X., Peng, C., & Fu, A. Z. (2003). Urbanization and health care in rural China. *Contemporary Economic Policy, 21*(1), 11–24.

Liu, Y., Yang, G., Zeng, Y., Horton, R., & Chen, L. (2013). Policy dialogue on China's changing burden of disease. *The Lancet, 381*(9882), 1961–1962.

Li, X., Wang, C., Zhang, G., Xiao, L., & Dixon, J. (2012). Urbanisation and human health in China: Spatial features and a systemic perspective. *Environmental Science and Pollution Research, 19*(5), 1375–1384.

Lukman, H., Dye, L., & Blundell, J. (1998). Relationship between diet and obesity in Chinese groups. *British Food Journal, 100*(1), 3–9.

McDade, T. W., & Adair, L. S. (2001). Defining the "urban" in urbanization and health: A factor analysis approach. *Social Science & Medicine, 53*(1), 55–70.

Miao, J., & Wu, X. (2016). Urbanization, socioeconomic status and health disparity in China. *Health & Place, 42*(Supplement C), 87–95.

Monda, K. L., Adair, L. S., Zhai, F., & Popkin, B. M. (2008). Longitudinal relationships between occupational and domestic physical activity patterns and body weight in China. *European Journal of Clinical Nutrition, 62*, 1318–1325.

Monda, K. L., Gordon-Larsen, P., Stevens, J., & Popkin, B. M. (2007). China's transition: The effect of rapid urbanization on adult occupational physical activity. *Social Science & Medicine, 64*(4), 858–870.

Moore, M., Gould, P., & Keary, B. S. (2003). Global urbanization and impact on health. *International Journal of Hygiene and Environmental Health, 206*(4), 269–278.

Mutatkar, R. (1995). Public health problems of urbanization. *Social Science & Medicine, 41*(7), 977–981.

Ng, M., Fleming, T., Robinson, M., Thomson, M., Graetz, N., Margono, C., et al. (2014). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: A systematic analysis for the global burden of disease study 2013. *The Lancet, 384*(9945), 766–781.

Ng, S. W., Norton, E. C., & Popkin, B. M. (2009). Why have physical activity levels declined among Chinese adults? Findings from the 1991–2006 China health and nutrition surveys. *Social Science & Medicine, 68*(7), 1305–1314.

van de Poel, E., O'Donnell, O., & van Doorslaer, E. (2012). Is there a health penalty of China's rapid urbanization? *Health Economics, 21*(4), 367–385.

Popkin, B. M. (2001). The nutrition transition and obesity in the developing world. *Journal of Nutrition, 131*(3), 871S–873S.

Popkin, B. M., & Du, S. (2003). Dynamics of the nutrition transition toward the animal foods sector in China and its implications: A worried perspective. *Journal of Nutrition, 133*(11), 3898S–3906S.

Popkin, B. M., Du, S., Zhai, F., & Zhang, B. (2010). Cohort profile: The China health and nutrition survey-monitoring and understanding socio-economic and health change in China, 1989-2011. *International Journal of Epidemiology, 39*(6), 14–35.

Popkin, B. M., & Gordon-Larsen, P. (2004). The nutrition transition: Worldwide obesity dynamics and their determinants. *International Journal of Obesity, 28*, S2–S9.

Popkin, B. M., Kim, S., Rusev, E., Du, S., & Zizza, C. (2006). Measuring the full economic costs of diet, physical activity and obesity-related chronic diseases. *Obesity Reviews, 7*(3), 271–293.

Reardon, T., & Berdegué, J. A. (2002). The rapid rise of supermarkets in Latin america: Challenges and opportunities for development. *Development Policy Review, 20*(4), 371–388.

Riley, N. E. (2004). China's population: New trends and challenges. *Population Bulletin, 59*(2), 1–36.

Sallis, J. F., Hsdkell, W. L., Wood, P. D., Fortmann, S. P., Rogers, T., Blair, S. N., et al. (1985). Physical activity assessment methodology in the five-city project. *American Journal of Epidemiology, 121*(1), 91–106.

Shao, M., Tang, X., Zhang, Y., & Li, W. (2006). City clusters in China: Air and surface water pollution. *Frontiers in Ecology and the Environment, 4*(7), 353–361.

Swinburn, B. A., Caterson, I. D., Seidell, J. C., & James, W. P. T. (2004). Diet, nutrition and the prevention of excess weight gain and obesity. *Public Health Nutrition, 7*(1A), 123–146.

Tian, H., Nan, Y., Hu, G., Dong, Q., Yang, X., Pietinen, P., et al. (1995). Dietary survey in a Chinese population. *European Journal of Clinical Nutrition, 49*(1), 26–32.

Tucker, L. A., & Kano, M. J. (1992). Dietary fat and body fat: A multivariate study of 205 adult females. *American Journal of Clinical Nutrition, 56*(4), 616–622.

United Nations (2014). *World urbanization prospects: The 2014 revision*. https://esa.un.org/unpd/wup/publications/files/wup2014-report.pdf, Accessed date: 4 February 2018.

Van de Poel, E., O'Donnell, O., & van Doorslaer, E. (2007). Are urban children really healthier? Evidence from 47 developing countries. *Social Science & Medicine, 65*(10), 1986–2003.

Van de Poel, E., O'Donnell, O., & van Doorslaer, E. (2009). Urbanization and the spread of diseases of affluence in China. *Economics and Human Biology, 7*(2), 200–216.
Villalonga-Olives, E., & Kawachi, I. (2017). The dark side of social capital: A systematic review of the negative health effects of social capital. *Social Science & Medicine, 194*, 105–127.
Vlahov, D., & Galea, S. (2002). Urbanization, urbanicity, and health. *Journal of Urban Health, 79*(1), S1–S12.
Wang, Y. P., Wang, Y., & Wu, J. (2009). Urbanization and informal development in China: Urban villages in shenzhen. *International Journal of Urban and Regional Research, 33*(4), 957–973.
World Health Organization (2013). *A global brief on hypertension: Silent killer, global public health crisis.* Geneva: WHO.
World Health Organization (2016). *Global report on urban health: Equitable, healthier cities for sustainable development*Geneva: WHO.
Yang, G., Wang, Y., Zeng, Y., Gao, G. F., Liang, X., Zhou, M., et al. (2013). Rapid health transition in China, 1990–2010: Findings from the global burden of disease study 2010. *The Lancet, 381*(9882), 1987–2015.
Yan, S., Li, J., Li, S., Zhang, B., Du, S., Gordon-Larsen, P., et al. (2012). The expanding burden of cardiometabolic risk in China: The China health and nutrition survey. *Obesity Reviews, 13*(9), 810–821.
Zhang, J., Mauzerall, D. L., Zhu, T., Liang, S., Ezzati, M., & Remais, J. V. (2010). Environmental health in China: Progress towards clean air and safe water. *The Lancet, 375*(9720), 1110–1119.
Zhu, Y. G., Ioannidis, J. P., Li, H., Jones, K. C., & Martin, F. L. (2011). Understanding and harnessing the health effects of rapid urbanization in China. *Environmental Science & Technology, 45*, 5099–5104.